GenCore version 5.1.6


M protein - protein search, using sw model

un on: November 21, 2003, 15:28:59 ; Search time 36 Seconds
(without alignments)
35.273 Million cell updates/sec

itle: US-10-064-903-1
erfect score: 29
equence: 1 HXXXHXXH 8

coring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

earched: 1107863 seqs, 158726573 residues

otal number of hits satisfying chosen parameters: 1107863

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| esult No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 23 | 79.3 | 30 | 22 | AAO13629 | Human polypeptide |
| 2 | 23 | 79.3 | 60 | 23 | ABP10050 | Human ORFX protein |
| 3 | 23 | 79.3 | 66 | 23 | ABP33078 | Human ORF2051 prot |
| 4 | 23 | 79.3 | 68 | 23 | ABP04899 | Human ORFX protein |
| 5 | 23 | 79.3 | 80 | 22 | AAU60888 | Propionibacterium |
| 6 | 23 | 79.3 | 92 | 22 | AAO11260 | Human polypeptide |
| 7 | 23 | 79.3 | 115 | 23 | ABP03095 | Human ORFX protein |
| 8 | 23 | 79.3 | 116 | 23 | ABP01907 | Human ORFX protein |
| 9 | 23 | 79.3 | 119 | 22 | AAM25686 | Human protein sequ |
| 10 | 23 | 79.3 | 132 | 22 | AAB63312 | Human breast cance |
| 11 | 23 | 79.3 | 134 | 22 | AAU67989 | Propionibacterium |
| 12 | 23 | 79.3 | 137 | 23 | ABP02968 | Human ORFX protein |
| 13 | 23 | 79.3 | 141 | 22 | AAO10851 | Human polypeptide |
| 14 | 23 | 79.3 | 167 | 22 | ABB70944 | Drosophila melanog |
| 15 | 23 | 79.3 | 218 | 23 | ABP31393 | Human ORF366 prote |
| 16 | 23 | 79.3 | 357 | 22 | AAU64020 | Propionibacterium |
| 17 | 23 | 79.3 | 462 | 22 | AAU42927 | Propionibacterium |
| 18 | 23 | 79.3 | 466 | 18 | AAW09825 | UDP-glucose:thiohy |
| 19 | 23 | 79.3 | 508 | 22 | ABB71345 | Drosophila melanog |
| 20 | 23 | 79.3 | 543 | 22 | ABG22945 | Novel human diagno |
| 21 | 23 | 79.3 | 572 | 24 | ABP77246 | N. gonorrhoeae ami |
| 22 | 23 | 79.3 | 635 | 18 | AAW19920 | Human Ksr' (kinase |
| 23 | 23 | 79.3 | 880 | 22 | ABB65766 | Drosophila melanog |
| 24 | 23 | 79.3 | 1078 | 24 | ABP96069 | Human protein kina |
| 25 | 23 | 79.3 | 1133 | 22 | ABB65544 | Drosophila melanog |
| 26 | 23 | 79.3 | 1187 | 22 | ABB67666 | Drosophila melanog |
| 27 | 23 | 79.3 | 1518 | 24 | ABJ18375 | Breast specific re |
| 28 | 23 | 79.3 | 1529 | 17 | AAR97985 | CORK potassium cha |
| 29 | 23 | 79.3 | 1575 | 22 | ABG27933 | Novel human diagno |
| 30 | 23 | 79.3 | 2424 | 22 | ABB58924 | Drosophila melanog |
| 31 | 23 | 79.3 | 3502 | 22 | ABB58382 | Drosophila melanog |
| 32 | 22 | 75.9 | 34 | 21 | AAG07703 | Arabidopsis thalia |
| 33 | 22 | 75.9 | 37 | 22 | AAM86640 | Human immune/haema |
| 34 | 22 | 75.9 | 41 | 21 | AAB34597 | Human secreted pro |
| 35 | 22 | 75.9 | 49 | 22 | ABB17194 | Human nervous syst |
| 36 | 22 | 75.9 | 50 | 23 | ABP31516 | Human ORF489 prote |
| 37 | 22 | 75.9 | 52 | 23 | ABG93190 | S. cerevisiae BAX- |
| 38 | 22 | 75.9 | 53 | 23 | ABP07950 | Human ORFX protein |
| 39 | 22 | 75.9 | 55 | 22 | AAU58208 | Propionibacterium |
| 40 | 22 | 75.9 | 56 | 23 | ABP32940 | Human ORF1913 prot |
| 41 | 22 | 75.9 | 57 | 22 | AAM78747 | Human protein SEQ |
| 42 | 22 | 75.9 | 59 | 22 | AAM79731 | Human protein SEQ |
| 43 | 22 | 75.9 | 60 | 23 | ABP05469 | Human ORFX protein |
| 44 | 22 | 75.9 | 61 | 22 | AAO10714 | Human polypeptide |
| 45 | 22 | 75.9 | 61 | 22 | AAG75865 | Human colon cancer |

ALIGNMENTS

RESULT 1
AAO13629
ID AAO13629 standard; Protein; 30 AA.
XX
AC AAO13629;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 27521.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
DR N-PSDB; AAI93560.

```
X
T   Isolated nucleic acids and polypeptides, useful for preventing
T   diagnosing and treating e.g. leukaemia, inflammation and immune
T   disorders -
X
S   Claim 20; SEQ ID NO 27521; 1399pp + Sequence Listing; English.
X
C   The invention relates to human polynucleotides (AAI79941-AAI93841) and
C   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
C   cytokine, cell proliferation or cell differentiation or which may induce
C   production of other cytokines in other cell populations. The
C   polynucleotides and polypeptides are useful in gene therapy, vaccines or
C   peptide therapy. The polypeptides have various cytokine-like activities,
C   e.g. stem cell growth factor activity, haematopoiesis regulating
C   activity, tissue growth factor activity, immunomodulatory activity and
C   activin/inhibin activity and may be useful in the diagnosis and/or
C   treatment of cancer, leukaemia, nervous system disorders, arthritis and
C   inflammation.
C   Note: The sequence data for this patent did not form part of the printed
C   specification, but was obtained in electronic format directly from WIPO
C   at ftp.wipo.int/pub/published_pct_sequences.
X
Q   Sequence   30 AA;

 Query Match             79.3%;  Score 23;  DB 22;  Length 30;
 Best Local Similarity   37.5%;  Pred. No. 4.3e+02;
 Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db         18 HTHTHTSH 25


RESULT 2
ABP10050
ID   ABP10050 standard; Protein; 60 AA.
XX
AC   ABP10050;
XX
DT   25-JUN-2002  (first entry)
XX
DE   Human ORFX protein sequence SEQ ID NO:20082.
XX
KW   Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW   hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW   degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW   cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW   hypertension; hypothyroidism; cholesterol ester storage disease;
KW   immune deficiency; immune disorder; infectious disease;
KW   autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW   myasthenia gravis.
XX
OS   Homo sapiens.
XX
PN   WO200192523-A2.
XX
PD   06-DEC-2001.
XX
PF   29-MAY-2001; 2001WO-US10836.
XX
PR   30-MAY-2000; 2000US-206132P.
PR   29-AUG-2000; 2000US-228716P.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach MD;
XX
DR   WPI; 2002-106308/14.
DR   N-PSDB; ABN25802.
XX
PT   Novel human polypeptides and polynucleotides useful for diagnosing,
PT   preventing and treating cardiovascular disease, neurodegenerative,
PT   hyperproliferative disorders and autoimmune disorders  -
XX
PS   Disclosure; SEQ ID 20082; 1037pp; English.
XX
CC   The present invention describes substantially purified human proteins
CC   (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC   in the specification). ABN15762 to ABN27252 encode the human ORFX
CC   proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC   treating or preventing a pathology associated with an ORFX-associated
CC   disorder in humans, and in the manufacture of a medicament for treating a
CC   syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC   sequences can be used in gene therapy. ORFX sequences can be used in the
CC   treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC   psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC   osteoarthritis, neurodegenerative disorders, disorders related to organ
CC   transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC   lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC   storage disease, various immune deficiencies and disorders, infectious
CC   diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC   arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC   disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC   useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC   bone degenerative disorders, or periodontal disease, and for gut
CC   protection or regeneration and treatment of lung or liver fibrosis,
CC   reperfusion injury in various tissues and conditions resulting from
CC   systemic cytokine damage.
CC   N.B.  The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   60 AA;

  Query Match             79.3%;  Score 23;  DB 23;  Length 60;
  Best Local Similarity   37.5%;  Pred. No. 7.4e+02;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db         20 HHSSHTTH 27


RESULT 3
ABP33078
ID   ABP33078 standard; Protein; 66 AA.
XX
AC   ABP33078;
XX
DT   09-JUL-2002  (first entry)
XX
DE   Human ORF2051 protein, SEQ ID NO:4102.
XX
KW   Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
KW   disease monitoring; cytokine; cell proliferation; cell differentiation;
KW   immune modulation; haematopoiesis regulation; tissue growth;
KW   angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; tumour inhibition; bodily characteristic; fertility;
KW   behaviour; cancer; proliferative disorder; neurological disorder;
KW   cardiovascular disease; immune system disorder; organ transplantation;
KW   tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
KW   hypothyroidism; cholesterol ester storage disease; infection; vulnerary;
KW   vasotropic; antipsoriatic; antidiabetic; cytostatic; nootropic;
KW   neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
KW   cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;
KW   dermatological; analgesic; virucide; antibacterial; fungicide.
XX
OS   Homo sapiens.
XX
PN   WO200190366-A2.
XX
PD   29-NOV-2001.
XX
PF   24-MAY-2001; 2001WO-US17076.
XX
PR   24-MAY-2000; 2000US-206690P.
```

```
X
A   (CURA-) CURAGEN CORP.
X
I   Leach MD,  Shimkets RA;
X
R   WPI; 2002-106200/14.
R   N-PSDB; ABN77104.
X
T   Novel human polypeptides and polynucleotides useful for diagnosing,
T   preventing and treating cardiovascular disease, neurodegenerative,
T   hyperproliferative disorders and disorders related to organ
T   transplantation  -
X
S   Claim 10; Page 1282-1283; 2508pp; English.
X
C   Sequences ABP31028-ABP35561 represent 4534 novel human proteins
C   designated ORF (open reading frame) 1-4534, and sequences ABN75054-
C   ABN79587 represent cDNAs encoding them. The invention also encompasses
C   polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
C   referred to as ORFX) proteins, polynucleotides at least 85% identical to
C   the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
C   polynucleotides, the recombinant production of ORFX proteins, antibodies
C   specific for ORFX proteins, methods of detecting ORFX polynucleotides and
C   polypeptides, methods of screening for modulators of ORFX expression or
C   activity, and methods of screening individuals for a predisposition to an
C   ORFX-associated disorder. The ORFX proteins of the invention have a wide
C   range of biological activities, such as cytokine, cell proliferation,
C   cell differentiation, immune modulation, haematopoiesis regulation,
C   tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
C   chemokinetic activity, haemostatic activity, thrombolytic activity,
C   receptor/ligand, antiinflammatory activity, tumour inhibition activity,
C   and antiinfective activity, and may also be involved in the determination
C   of bodily characteristics, fertility and behaviour. ORFX proteins,
C   nucleic acids and antibodies may be used in the treatment of cancers,
C   other proliferative disorders such as psoriasis and benign tumours,
C   neurological disorders such as epilepsy and Alzheimer's disease,
C   cardiovascular diseases, immune system disorders, disorders related to
C   organ transplantation, disorders of tissue growth and regeneration,
C   diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
C   storage disease, and infectious diseases caused by viral, bacterial,
C   fungal and other pathogens. ORFX nucleic acids may also be used as a
C   source of primers and probes, in the detection of ORFX genomic sequences
C   or transcripts, in the identification and cloning of homologous
C   sequences, in genetic diagnosis, and in forensic biology. The ORFX
C   nucleic acids may additionally be used to produce transgenic animals
C   which may be useful for studying the function and/or activity of ORFX
C   protein, and in drug screening. The ORFX proteins may also be used as
C   immunogens to generate specific antibodies, which are useful in the
C   diagnosis, treatment and monitoring of ORFX-associated diseases.
X
Q   Sequence   66 AA;

 Query Match             79.3%;  Score 23;  DB 23;  Length 66;
 Best Local Similarity   37.5%;  Pred. No. 8e+02;
 Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

y          1 HXXXHXXH 8
             |  |  |
b         28 HHTTHTSH 35


ESULT 4
BP04899
D   ABP04899 standard; Protein; 68 AA.
X
C   ABP04899;
X
T   25-JUN-2002  (first entry)
X
E   Human ORFX protein sequence SEQ ID NO:9780.
X
W   Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
W   hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW   degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW   cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW   hypertension; hypothyroidism; cholesterol ester storage disease;
KW   immune deficiency; immune disorder; infectious disease;
KW   autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW   myasthenia gravis.
XX
OS   Homo sapiens.
XX
PN   WO200192523-A2.
XX
PD   06-DEC-2001.
XX
PF   29-MAY-2001; 2001WO-US10836.
XX
PR   30-MAY-2000; 2000US-206132P.
PR   29-AUG-2000; 2000US-228716P.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach MD;
XX
DR   WPI; 2002-106308/14.
DR   N-PSDB; ABN20651.
XX
PT   Novel human polypeptides and polynucleotides useful for diagnosing,
PT   preventing and treating cardiovascular disease, neurodegenerative,
PT   hyperproliferative disorders and autoimmune disorders  -
XX
PS   Disclosure; SEQ ID 9780; 1037pp; English.
XX
CC   The present invention describes substantially purified human proteins
CC   (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC   in the specification). ABN15762 to ABN27252 encode the human ORFX
CC   proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC   treating or preventing a pathology associated with an ORFX-associated
CC   disorder in humans, and in the manufacture of a medicament for treating a
CC   syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC   sequences can be used in gene therapy. ORFX sequences can be used in the
CC   treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC   psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC   osteoarthritis, neurodegenerative disorders, disorders related to organ
CC   transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC   lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC   storage disease, various immune deficiencies and disorders, infectious
CC   diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC   arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC   disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC   useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC   bone degenerative disorders, or periodontal disease, and for gut
CC   protection or regeneration and treatment of lung or liver fibrosis,
CC   reperfusion injury in various tissues and conditions resulting from
CC   systemic cytokine damage.
CC   N.B.  The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   68 AA;

  Query Match             79.3%;  Score 23;  DB 23;  Length 68;
  Best Local Similarity   37.5%;  Pred. No. 8.2e+02;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |  |  |
Db          9 HTHSHTAH 16


RESULT 5
AAU60888
ID   AAU60888 standard; Protein; 80 AA.
XX
AC   AAU60888;
```

XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #21784.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
DR N-PSDB; AAS59613.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
PS Example 1; SEQ ID No 22083; 1069pp; English.
XX
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 80 AA;

Query Match 79.3%; Score 23; DB 22; Length 80;
Best Local Similarity 37.5%; Pred. No. 9.4e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 HXXXHXXH 8
Db 47 HSASHRTH 54

RESULT 6
AO11260
ID AAO11260 standard; Protein; 92 AA.
XX
AC AAO11260;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 25152.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
DR N-PSDB; AAI91191.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
PS Claim 20; SEQ ID NO 25152; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 92 AA;

Query Match 79.3%; Score 23; DB 22; Length 92;
Best Local Similarity 37.5%; Pred. No. 1e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 HXXXHXXH 8
Db 41 HTTAHGSH 48

RESULT 7
ABP03095
ID ABP03095 standard; Protein; 115 AA.
XX
AC ABP03095;
XX
DT 24-JUN-2002 (first entry)
XX
DE Human ORFX protein sequence SEQ ID NO:6172.
XX
KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;

```
W   cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
W   hypertension; hypothyroidism; cholesterol ester storage disease;
W   immune deficiency; immune disorder; infectious disease;
W   autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
W   myasthenia gravis.
X
S   Homo sapiens.
X
N   WO200192523-A2.
X
D   06-DEC-2001.
X
F   29-MAY-2001; 2001WO-US10836.
X
R   30-MAY-2000; 2000US-206132P.
R   29-AUG-2000; 2000US-228716P.
X
A   (CURA-) CURAGEN CORP.
X
I   Shimkets RA,  Leach MD;
X
R   WPI; 2002-106308/14.
R   N-PSDB; ABN18847.
X
T   Novel human polypeptides and polynucleotides useful for diagnosing,
T   preventing and treating cardiovascular disease, neurodegenerative,
T   hyperproliferative disorders and autoimmune disorders  -
X
S   Disclosure; SEQ ID 6172; 1037pp; English.
X
C   The present invention describes substantially purified human proteins
C   (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
C   in the specification). ABN15762 to ABN27252 encode the human ORFX
C   proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
C   treating or preventing a pathology associated with an ORFX-associated
C   disorder in humans, and in the manufacture of a medicament for treating a
C   syndrome associated with ORFX-associated disorder. ORFX polynucleotide
C   sequences can be used in gene therapy. ORFX sequences can be used in the
C   treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
C   psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
C   osteoarthritis, neurodegenerative disorders, disorders related to organ
C   transplantation, cardiovascular diseases, diabetes mellitus, systemic
C   lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
C   storage disease, various immune deficiencies and disorders, infectious
C   diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
C   arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
C   disease and autoimmune inflammatory eye disease. ORFX proteins are also
C   useful for treating burns, incisions, ulcers, for treating osteoporosis,
C   bone degenerative disorders, or periodontal disease, and for gut
C   protection or regeneration and treatment of lung or liver fibrosis,
C   reperfusion injury in various tissues and conditions resulting from
C   systemic cytokine damage.
C   N.B.  The sequence data for this patent did not form part of the printed
C   specification, but was obtained in electronic format directly from WIPO
C   at ftp.wipo.int/pub/published_pct_sequences.
X
Q   Sequence   115 AA;

 Query Match             79.3%;  Score 23;  DB 23;  Length 115;
 Best Local Similarity   37.5%;  Pred. No. 1.3e+03;
 Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

y          1 HXXXHXXH 8
             |   |  |
b         45 HTTTHTLH 52


ESULT 8
BP01907
D   ABP01907 standard; Protein; 116 AA.
X
C   ABP01907;
X
DT   24-JUN-2002  (first entry)
XX
DE   Human ORFX protein sequence SEQ ID NO:3796.
XX
KW   Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW   hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW   degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW   cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW   hypertension; hypothyroidism; cholesterol ester storage disease;
KW   immune deficiency; immune disorder; infectious disease;
KW   autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW   myasthenia gravis.
XX
OS   Homo sapiens.
XX
PN   WO200192523-A2.
XX
PD   06-DEC-2001.
XX
PF   29-MAY-2001; 2001WO-US10836.
XX
PR   30-MAY-2000; 2000US-206132P.
PR   29-AUG-2000; 2000US-228716P.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach MD;
XX
DR   WPI; 2002-106308/14.
DR   N-PSDB; ABN17659.
XX
PT   Novel human polypeptides and polynucleotides useful for diagnosing,
PT   preventing and treating cardiovascular disease, neurodegenerative,
PT   hyperproliferative disorders and autoimmune disorders  -
XX
PS   Disclosure; SEQ ID 3796; 1037pp; English.
XX
CC   The present invention describes substantially purified human proteins
CC   (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC   in the specification). ABN15762 to ABN27252 encode the human ORFX
CC   proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC   treating or preventing a pathology associated with an ORFX-associated
CC   disorder in humans, and in the manufacture of a medicament for treating a
CC   syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC   sequences can be used in gene therapy. ORFX sequences can be used in the
CC   treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC   psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC   osteoarthritis, neurodegenerative disorders, disorders related to organ
CC   transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC   lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC   storage disease, various immune deficiencies and disorders, infectious
CC   diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC   arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC   disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC   useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC   bone degenerative disorders, or periodontal disease, and for gut
CC   protection or regeneration and treatment of lung or liver fibrosis,
CC   reperfusion injury in various tissues and conditions resulting from
CC   systemic cytokine damage.
CC   N.B.  The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   116 AA;

  Query Match             79.3%;  Score 23;  DB 23;  Length 116;
  Best Local Similarity   37.5%;  Pred. No. 1.3e+03;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db         46 HSTAHSHH 53
```

```
ESULT 9
AM25686
D   AAM25686 standard; Protein; 119 AA.
X
C   AAM25686;
X
T   16-OCT-2001  (first entry)
X
E   Human protein sequence SEQ ID NO:1201.
X
W   Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
W   antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
W   antibacterial; endocrine; cardiant; central nervous system; virucide;
W   anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
W   antiaggregant; haemostatic; vulnerary; antiulcer; osteopathic; eczema;
W   dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
W   neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
W   immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
W   antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
W   cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
W   genetic disease; haematopoietic disorder; platelet disorder; asthma;
W   thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
W   allergic rhinitis; diabetes; multiple sclerosis; depression;
W   Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
W   neurological disorder.
X
S   Homo sapiens.
X
N   WO200153455-A2.
X
D   26-JUL-2001.
X
F   22-DEC-2000; 2000WO-US35017.
X
R   23-DEC-1999;   99US-0471275.
R   21-JAN-2000; 2000US-0488725.
R   25-APR-2000; 2000US-0552317.
X
A   (HYSE-) HYSEQ INC.
X
I   Tang YT,  Liu C,  Drmanac RT;
X
R   WPI; 2001-457603/49.
R   N-PSDB; AAH99627.
X
T   Isolated human polynucleotides encoding polypeptides, useful for the
T   treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
X
S   Claim 20; Page 247; 1217pp; English.
X
C   AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
C   AAM25963. The proteins can have activities based on the tissues and
C   cells they are expressed in, such as: antiinflammatory; antirheumatic;
C   antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
C   central nervous system; virucide; anti-HIV; fungicide; antimutagen;
C   cardiovascular; antianaemic; antiaggregant; haemostatic; vulnerary;
C   antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;
C   antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
C   antiparkinsonian; and immunostimulant. The proteins and polynucleotides
C   encoding them can be used in gene therapy, antisense therapy and vaccine
C   production, The proteins and polynucleotides are useful for screening for
C   agonists or antagonists of a protein and for the treatment and diagnosis
C   of disorders associated with the activity of a protein e.g. inflammation,
C   rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
C   neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
C   infections, autoimmunity, genetic diseases, haematopoietic disorders,
C   anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
C   osteoporosis, severe combined immunodeficiency, eczema, allergic
C   rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
C   Alzheimer's disease, Parkinson's disease, neurodegenerative and
C   neurological disorders.
X
```

```
SQ   Sequence   119 AA;

  Query Match             79.3%;  Score 23;  DB 22;  Length 119;
  Best Local Similarity   37.5%;  Pred. No. 1.3e+03;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db          7 HASAHSGH 14


RESULT 10
AAB63312
ID   AAB63312 standard; Protein; 132 AA.
XX
AC   AAB63312;
XX
DT   26-MAR-2001  (first entry)
XX
DE   Human breast cancer associated antigen protein sequence SEQ ID NO:674.
XX
KW   Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW   cancer associated antigen; cytostatic; cancer vaccine.
XX
OS   Homo sapiens.
XX
PN   WO200073801-A2.
XX
PD   07-DEC-2000.
XX
PF   26-MAY-2000; 2000WO-US14749.
XX
PR   28-MAY-1999;   99US-0136526.
PR   10-SEP-1999;   99US-0153454.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Obata Y;
XX
DR   WPI; 2001-025274/03.
XX
PT   Nucleic acids  encoding breast, gastric and prostate cancer associated
PT   antigen precursors, useful for diagnosing and treating a condition
PT   characterized by expression of an abnormal amount of a protein, e.g.
PT   cancer -
XX
PS   Example 1; Page 512; 799pp; English.
XX
CC   AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
CC   represent nucleotide sequences encoding human breast, gastric and
CC   prostate cancer associated antigen precursors (CAAP) respectively.
CC   AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
CC   represent human breast, gastric and prostate CAAP protein sequence
CC   respectively. CAAPs have cytostatic activity and can be used in the
CC   production of cancer vaccines. The human CAAP proteins, peptides, nucleic
CC   acids or anti-CAAP antibodies are useful for diagnosing and treating a
CC   condition characterised by expression of an abnormal amount of a protein,
CC   e.g. cancer.
XX
SQ   Sequence   132 AA;

  Query Match             79.3%;  Score 23;  DB 22;  Length 132;
  Best Local Similarity   37.5%;  Pred. No. 1.4e+03;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db         98 HSHTHTTH 105


RESULT 11
AAU67989
ID   AAU67989 standard; Protein; 134 AA.
```

AAU67989;

27-FEB-2002 (first entry)

Propionibacterium acnes immunogenic protein #28885.

SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis; uveitis; endophthalmitis; bone; joint; central nervous system; ELISA; inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay; dermatological; osteopathic; neuroprotectant.

Propionibacterium acnes.

WO200181581-A2.

01-NOV-2001.

20-APR-2001; 2001WO-US12865.

21-APR-2000; 2000US-199047P.
02-JUN-2000; 2000US-208841P.
07-JUL-2000; 2000US-216747P.

(CORI-) CORIXA CORP.

Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
L'maisonneuve J, Zhang Y, Jen S, Carter D;

WPI; 2001-616774/71.
N-PSDB; AAS59785.

Propionibacterium acnes polypeptides and nucleic acids useful for vaccinating against and diagnosing infections, especially useful for treating acne vulgaris -

Example 1; SEQ ID No 29184; 1069pp; English.

Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic polypeptides. The proteins and their associated DNA sequences are used in the treatment, prevention and diagnosis of medical conditions caused by P. acnes. The disorders include SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved in infections of bone, joints and the central nervous system, however it is particularly involved in the inflammatory lesions associated with acne vulgaris. A method for detecting the presence or absence of P. acnes in a patient comprises contacting a sample with a binding agent that binds to the proteins of the invention and determining the amount of bound protein in the sample. The polypeptides may be used as antigens in the production of antibodies specific for P. acnes proteins. These antibodies can be used to downregulate expression and activity of P. acnes polypeptides and therefore treat P. acnes infections. The antibodies may also be used as diagnostic agents for determining P. acnes presence, for example, by enzyme linked immunosorbent assay (ELISA).
Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 134 AA;

```
 Query Match             79.3%;  Score 23;  DB 22;  Length 134;
 Best Local Similarity   37.5%;  Pred. No. 1.4e+03;
 Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Y          1 HXXXHXXH 8
             |   |  |
o         40 HSTAHSHH 47
```

ESULT 12
SP02968
D ABP02968 standard; Protein; 137 AA.

XX
AC ABP02968;
XX
DT 25-JUN-2002 (first entry)
XX
DE Human ORFX protein sequence SEQ ID NO:5918.
XX
KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis.
XX
OS Homo sapiens.
XX
PN WO200192523-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US10836.
XX
PR 30-MAY-2000; 2000US-206132P.
PR 29-AUG-2000; 2000US-228716P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach MD;
XX
DR WPI; 2002-106308/14.
DR N-PSDB; ABN18720.
XX
PT Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders -
XX
PS Disclosure; SEQ ID 5918; 1037pp; English.
XX
CC The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC in the specification). ABN15762 to ABN27252 encode the human ORFX
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC disorder in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, disorders related to organ
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 137 AA;

```
  Query Match             79.3%;  Score 23;  DB 23;  Length 137;
  Best Local Similarity   37.5%;  Pred. No. 1.4e+03;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
```

```
                |   |  |
>         77 HTTHHTTH 84


ESULT 13
AO10851
>   AAO10851 standard; Protein; 141 AA.
X
C   AAO10851;
X
T   06-NOV-2001  (first entry)
X
E   Human polypeptide SEQ ID NO 24743.
X
K   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
K   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
K   tissue growth factor; immunomodulatory; cancer; leukaemia;
K   nervous system disorders; arthritis; inflammation.
X
S   Homo sapiens.
X
N   WO200164835-A2.
X
D   07-SEP-2001.
X
F   26-FEB-2001; 2001WO-US04927.
X
R   28-FEB-2000; 2000US-0515126.
R   18-MAY-2000; 2000US-0577409.
X
A   (HYSE-) HYSEQ INC.
X
I   Tang YT,  Liu C,  Drmanac RT;
X
R   WPI; 2001-514838/56.
R   N-PSDB; AAI90782.
X
T   Isolated nucleic acids and polypeptides, useful for preventing
T   diagnosing and treating e.g. leukaemia, inflammation and immune
T   disorders -
X
S   Claim 20; SEQ ID NO 24743; 1399pp + Sequence Listing; English.
X
C   The invention relates to human polynucleotides (AAI79941-AAI93841) and
C   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
C   cytokine, cell proliferation or cell differentiation or which may induce
C   production of other cytokines in other cell populations. The
C   polynucleotides and polypeptides are useful in gene therapy, vaccines or
C   peptide therapy. The polypeptides have various cytokine-like activities,
C   e.g. stem cell growth factor activity, haematopoiesis regulating
C   activity, tissue growth factor activity, immunomodulatory activity and
C   activin/inhibin activity and may be useful in the diagnosis and/or
C   treatment of cancer, leukaemia, nervous system disorders, arthritis and
C   inflammation.
C   Note: The sequence data for this patent did not form part of the printed
C   specification, but was obtained in electronic format directly from WIPO
C   at ftp.wipo.int/pub/published_pct_sequences.
X
Q   Sequence   141 AA;

 Query Match             79.3%;  Score 23;  DB 22;  Length 141;
 Best Local Similarity   37.5%;  Pred. No. 1.5e+03;
 Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Y          1 HXXXHXXH 8
             |   |  |
b          8 HASAHAYH 15


ESULT 14
BB70944
D   ABB70944 standard; Protein; 167 AA.
X
AC   ABB70944;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 39624.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL15047.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 39624; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   167 AA;

  Query Match             79.3%;  Score 23;  DB 22;  Length 167;
  Best Local Similarity   37.5%;  Pred. No. 1.7e+03;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db        113 HSHSHAAH 120


RESULT 15
ABP31393
ID   ABP31393 standard; Protein; 218 AA.
XX
AC   ABP31393;
XX
DT   08-JUL-2002  (first entry)
XX
DE   Human ORF366 protein, SEQ ID NO:732.
XX
KW   Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
KW   disease monitoring; cytokine; cell proliferation; cell differentiation;
KW   immune modulation; haematopoiesis regulation; tissue growth;
KW   angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; tumour inhibition; bodily characteristic; fertility;
KW   behaviour; cancer; proliferative disorder; neurological disorder;
KW   cardiovascular disease; immune system disorder; organ transplantation;
```

W tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
W hypothyroidism; cholesterol ester storage disease; infection; vulnerary;
W vasotropic; antipsoriatic; antidiabetic; cytostatic; nootropic;
W neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
W cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;
W dermatological; analgesic; virucide; antibacterial; fungicide.
X
S Homo sapiens.
X
N WO200190366-A2.
X
D 29-NOV-2001.
X
F 24-MAY-2001; 2001WO-US17076.
X
R 24-MAY-2000; 2000US-206690P.
X
A (CURA-) CURAGEN CORP.
X
I Leach MD, Shimkets RA;
X
R WPI; 2002-106200/14.
R N-PSDB; ABN75419.
X
T Novel human polypeptides and polynucleotides useful for diagnosing,
T preventing and treating cardiovascular disease, neurodegenerative,
T hyperproliferative disorders and disorders related to organ
T transplantation -
X
S Claim 10; Page 450; 2508pp; English.
X
C Sequences ABP31028-ABP35561 represent 4534 novel human proteins
C designated ORF (open reading frame) 1-4534, and sequences ABN75054-
C ABN79587 represent cDNAs encoding them. The invention also encompasses
C polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
C referred to as ORFX) proteins, polynucleotides at least 85% identical to
C the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
C polynucleotides, the recombinant production of ORFX proteins, antibodies
C specific for ORFX proteins, methods of detecting ORFX polynucleotides and
C polypeptides, methods of screening for modulators of ORFX expression or
C activity, and methods of screening individuals for a predisposition to an
C ORFX-associated disorder. The ORFX proteins of the invention have a wide
C range of biological activities, such as cytokine, cell proliferation,
C cell differentiation, immune modulation, haematopoiesis regulation,
C tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
C chemokinetic activity, haemostatic activity, thrombolytic activity,
C receptor/ligand, antiinflammatory activity, tumour inhibition activity,
C and antiinfective activity, and may also be involved in the determination
C of bodily characteristics, fertility and behaviour. ORFX proteins,
C nucleic acids and antibodies may be used in the treatment of cancers,
C other proliferative disorders such as psoriasis and benign tumours,
C neurological disorders such as epilepsy and Alzheimer's disease,
C cardiovascular diseases, immune system disorders, disorders related to
C organ transplantation, disorders of tissue growth and regeneration,
C diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
C storage disease, and infectious diseases caused by viral, bacterial,
C fungal and other pathogens. ORFX nucleic acids may also be used as a
C source of primers and probes, in the detection of ORFX genomic sequences
C or transcripts, in the identification and cloning of homologous
C sequences, in genetic diagnosis, and in forensic biology. The ORFX
C nucleic acids may additionally be used to produce transgenic animals
C which may be useful for studying the function and/or activity of ORFX
C protein, and in drug screening. The ORFX proteins may also be used as
C immunogens to generate specific antibodies, which are useful in the
C diagnosis, treatment and monitoring of ORFX-associated diseases.
X
Q Sequence 218 AA;

Query Match 79.3%; Score 23; DB 23; Length 218;
Best Local Similarity 37.5%; Pred. No. 2.1e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Y 1 HXXXHXXH 8

Db 86 HTHSHAAH 93

Search completed: November 21, 2003, 15:48:01
Job time : 37 secs

GenCore version 5.1.6


M protein - protein search, using sw model

un on: November 21, 2003, 15:46:45 ; Search time 14.5 Seconds
(without alignments)
23.344 Million cell updates/sec

itle: US-10-064-903-1
erfect score: 29
equence: 1 HXXXHXXH 8

coring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

earched: 328717 seqs, 42310858 residues

otal number of hits satisfying chosen parameters: 328717

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| esult No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 23 | 79.3 | 139 | 4 | US-09-252-991A-32472 | Sequence 32472, A |
| 2 | 23 | 79.3 | 511 | 4 | US-09-252-991A-22789 | Sequence 22789, A |
| 3 | 23 | 79.3 | 573 | 4 | US-09-252-991A-24488 | Sequence 24488, A |
| 4 | 23 | 79.3 | 635 | 1 | US-08-571-758-10 | Sequence 10, Appl |
| 5 | 23 | 79.3 | 635 | 1 | US-08-909-984A-10 | Sequence 10, Appl |
| 6 | 23 | 79.3 | 635 | 1 | US-08-909-983-10 | Sequence 10, Appl |
| 7 | 22 | 75.9 | 151 | 2 | US-08-858-767-30 | Sequence 30, Appl |
| 8 | 22 | 75.9 | 151 | 2 | US-08-863-028-30 | Sequence 30, Appl |
| 9 | 22 | 75.9 | 179 | 4 | US-09-615-192A-289 | Sequence 289, App |
| 10 | 22 | 75.9 | 249 | 4 | US-09-252-991A-22610 | Sequence 22610, A |
| 11 | 22 | 75.9 | 260 | 4 | US-09-252-991A-20987 | Sequence 20987, A |
| 12 | 22 | 75.9 | 272 | 4 | US-08-858-207A-447 | Sequence 447, App |
| 13 | 22 | 75.9 | 323 | 4 | US-09-328-352-6181 | Sequence 6181, Ap |
| 14 | 22 | 75.9 | 355 | 2 | US-08-758-621-4 | Sequence 4, Appli |
| 15 | 22 | 75.9 | 355 | 3 | US-09-107-858-4 | Sequence 4, Appli |
| 16 | 22 | 75.9 | 369 | 4 | US-09-252-991A-25533 | Sequence 25533, A |
| 17 | 22 | 75.9 | 387 | 4 | US-09-364-230-18 | Sequence 18, Appl |
| 18 | 22 | 75.9 | 388 | 4 | US-09-252-991A-31265 | Sequence 31265, A |
| 19 | 22 | 75.9 | 412 | 4 | US-09-252-991A-26532 | Sequence 26532, A |
| 20 | 22 | 75.9 | 413 | 4 | US-09-328-352-5589 | Sequence 5589, Ap |
| 21 | 22 | 75.9 | 431 | 1 | US-08-311-023-2 | Sequence 2, Appli |
| 22 | 22 | 75.9 | 447 | 4 | US-09-252-991A-32122 | Sequence 32122, A |
| 23 | 22 | 75.9 | 481 | 4 | US-09-252-991A-24508 | Sequence 24508, A |
| 24 | 22 | 75.9 | 504 | 4 | US-09-252-991A-28224 | Sequence 28224, A |
| 25 | 22 | 75.9 | 516 | 4 | US-09-252-991A-29719 | Sequence 29719, A |
| 26 | 22 | 75.9 | 533 | 4 | US-09-252-991A-23560 | Sequence 23560, A |
| 27 | 22 | 75.9 | 542 | 4 | US-09-107-532A-4858 | Sequence 4858, Ap |
| 28 | 22 | 75.9 | 559 | 2 | US-08-756-317-7 | Sequence 7, Appli |
| 29 | 22 | 75.9 | 559 | 2 | US-08-756-317-10 | Sequence 10, Appl |
| 30 | 22 | 75.9 | 559 | 4 | US-09-672-749-2 | Sequence 2, Appli |
| 31 | 22 | 75.9 | 559 | 4 | US-09-821-016-1 | Sequence 1, Appli |
| 32 | 22 | 75.9 | 582 | 4 | US-09-252-991A-27626 | Sequence 27626, A |
| 33 | 22 | 75.9 | 582 | 4 | US-09-252-991A-32678 | Sequence 32678, A |
| 34 | 22 | 75.9 | 637 | 4 | US-09-252-991A-28952 | Sequence 28952, A |
| 35 | 22 | 75.9 | 706 | 4 | US-09-252-991A-25730 | Sequence 25730, A |
| 36 | 22 | 75.9 | 795 | 4 | US-09-193-562D-11 | Sequence 11, Appl |
| 37 | 22 | 75.9 | 821 | 4 | US-09-193-562D-12 | Sequence 12, Appl |
| 38 | 22 | 75.9 | 834 | 4 | US-09-187-999-11 | Sequence 11, Appl |
| 39 | 22 | 75.9 | 872 | 2 | US-08-844-057-2 | Sequence 2, Appli |
| 40 | 22 | 75.9 | 872 | 4 | US-09-006-730-2 | Sequence 2, Appli |
| 41 | 22 | 75.9 | 876 | 1 | US-08-785-071A-2 | Sequence 2, Appli |
| 42 | 22 | 75.9 | 876 | 3 | US-09-012-872-2 | Sequence 2, Appli |
| 43 | 22 | 75.9 | 893 | 4 | US-09-328-352-6626 | Sequence 6626, Ap |
| 44 | 22 | 75.9 | 897 | 4 | US-09-134-001C-3600 | Sequence 3600, Ap |
| 45 | 22 | 75.9 | 905 | 4 | US-09-193-562D-2 | Sequence 2, Appli |

ALIGNMENTS

RESULT 1
US-09-252-991A-32472
; Sequence 32472, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32472
; LENGTH: 139
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32472

Query Match 79.3%; Score 23; DB 4; Length 139;
Best Local Similarity 37.5%; Pred. No. 5.7e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy          1 HXXXHXXH 8
              |   | | |
Db         26 HTALHSSH 33
```

RESULT 2
US-09-252-991A-22789
; Sequence 22789, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22789
; LENGTH: 511
; TYPE: PRT

```
   ORGANISM: Pseudomonas aeruginosa
S-09-252-991A-22789

 Query Match             79.3%;  Score 23;  DB 4;  Length 511;
 Best Local Similarity   37.5%;  Pred. No. 1.5e+03;
 Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

y          1 HXXXHXXH 8
             |  |  |
b        101 HAAAHAFH 108


ESULT 3
S-09-252-991A-24488
 Sequence 24488, Application US/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:
  APPLICANT:  Marc J. Rubenfield et al.
  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
  FILE REFERENCE:  107196.136
  CURRENT APPLICATION NUMBER: US/09/252,991A
  CURRENT FILING DATE:  1999-02-18
  PRIOR APPLICATION NUMBER:  US 60/074,788
  PRIOR FILING DATE:  1998-02-18
  PRIOR APPLICATION NUMBER:  US 60/094,190
  PRIOR FILING DATE:  1998-07-27
  NUMBER OF SEQ ID NOS:  33142
 SEQ ID NO 24488
   LENGTH: 573
   TYPE: PRT
   ORGANISM: Pseudomonas aeruginosa
'S-09-252-991A-24488

 Query Match             79.3%;  Score 23;  DB 4;  Length 573;
 Best Local Similarity   37.5%;  Pred. No. 1.7e+03;
 Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

)y          1 HXXXHXXH 8
              |  |  |
)b         18 HHAAHAAH 25


:ESULT 4
'S-08-571-758-10
 Sequence 10, Application US/08571758
 Patent No. 5700675
  GENERAL INFORMATION:
    APPLICANT:  Rubin, Gerry M.
    APPLICANT:  Therrien, Marc
    APPLICANT:  Chang, Henry C.
    APPLICANT:  Karim, Felix D.
    APPLICANT:  Wassarman, David A.
    TITLE OF INVENTION:  A No. 5700675el Protein Kinase Required for Ras
    TITLE OF INVENTION:  Signal Transduction
    NUMBER OF SEQUENCES:  12
    CORRESPONDENCE ADDRESS:
      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
      STREET:  268 BUSH STREET, SUITE 3200
      CITY:  SAN FRANCISCO
      STATE:  CALIFORNIA
      COUNTRY:  USA
      ZIP:  94104
    COMPUTER READABLE FORM:
      MEDIUM TYPE:  Floppy disk
      COMPUTER:  IBM PC compatible
      OPERATING SYSTEM:  PC-DOS/MS-DOS
      SOFTWARE:  PatentIn Release #1.0, Version #1.30
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER:  US/08/571,758
      FILING DATE:
      CLASSIFICATION:  435
    ATTORNEY/AGENT INFORMATION:
;      NAME:  OSMAN, RICHARD A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  B96-010
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 343-4341
;      TELEFAX:  (415) 343-4342
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  635 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-08-571-758-10

  Query Match             79.3%;  Score 23;  DB 1;  Length 635;
  Best Local Similarity   37.5%;  Pred. No. 1.8e+03;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |  |  |
Db         17 HTSAHTQH 24


RESULT 5
US-08-909-984A-10
; Sequence 10, Application US/08909984A
; Patent No. 5747275
;  GENERAL INFORMATION:
;    APPLICANT:  Rubin, Gerry M.
;    APPLICANT:  Therrien, Marc
;    APPLICANT:  Chang, Henry C.
;    APPLICANT:  Karim, Felix D.
;    APPLICANT:  Wassarman, David A.
;    TITLE OF INVENTION:  A No. 5747275el Protein Kinase Required for Ras
;    TITLE OF INVENTION:  Signal Transduction
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
;      STREET:  268 BUSH STREET, SUITE 3200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/909,984A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OSMAN, RICHARD A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  B96-010
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 343-4341
;      TELEFAX:  (415) 343-4342
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  635 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-08-909-984A-10

  Query Match             79.3%;  Score 23;  DB 1;  Length 635;
  Best Local Similarity   37.5%;  Pred. No. 1.8e+03;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;
```

```
Qy          1 HXXXHXXH 8
              |   |  |
Db         17 HTSAHTQH 24


RESULT 6
S-08-909-983-10
 Sequence 10, Application US/08909983
 Patent No. 5747288
  GENERAL INFORMATION:
    APPLICANT:  Rubin, Gerry M.
    APPLICANT:  Therrien, Marc
    APPLICANT:  Chang, Henry C.
    APPLICANT:  Karim, Felix D.
    APPLICANT:  Wassarman, David A.
    TITLE OF INVENTION:  A No. 5747288el Protein Kinase Required for Ras
    TITLE OF INVENTION:  Signal Transduction
    NUMBER OF SEQUENCES:  12
    CORRESPONDENCE ADDRESS:
      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
      STREET:  268 BUSH STREET, SUITE 3200
      CITY:  SAN FRANCISCO
      STATE:  CALIFORNIA
      COUNTRY:  USA
      ZIP:  94104
    COMPUTER READABLE FORM:
      MEDIUM TYPE:  Floppy disk
      COMPUTER:  IBM PC compatible
      OPERATING SYSTEM:  PC-DOS/MS-DOS
      SOFTWARE:  PatentIn Release #1.0, Version #1.30
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER:  US/08/909,983
      FILING DATE:  12-JUN-1997
      CLASSIFICATION:  435
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  08/571,758
      FILING DATE:
    ATTORNEY/AGENT INFORMATION:
      NAME:  OSMAN, RICHARD A
      REGISTRATION NUMBER:  36,627
      REFERENCE/DOCKET NUMBER:  B96-010
    TELECOMMUNICATION INFORMATION:
      TELEPHONE:  (415) 343-4341
      TELEFAX:  (415) 343-4342
  INFORMATION FOR SEQ ID NO:  10:
    SEQUENCE CHARACTERISTICS:
      LENGTH:  635 amino acids
      TYPE:  amino acid
      STRANDEDNESS:  not relevant
      TOPOLOGY:  not relevant
    MOLECULE TYPE:  peptide
S-08-909-983-10

 Query Match             79.3%;  Score 23;  DB 1;  Length 635;
 Best Local Similarity   37.5%;  Pred. No. 1.8e+03;
 Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db         17 HTSAHTQH 24


RESULT 7
S-08-858-767-30
 Sequence 30, Application US/08858767
 Patent No. 5837468
  GENERAL INFORMATION:
    APPLICANT:  WANG, Xun
    APPLICANT:  DUVICK, Jonathan P.
    APPLICANT:  BRIGGS, Steven P.
    TITLE OF INVENTION:  PCR-BASED CDNA SUBTRACTIVE CLONING
    TITLE OF INVENTION:  METHOD
    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/858,767
;      FILING DATE:  19-MAY-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/481,687
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  33229/325/PIHI
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  151 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-858-767-30

  Query Match             75.9%;  Score 22;  DB 2;  Length 151;
  Best Local Similarity   37.5%;  Pred. No. 8.9e+02;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db         56 HAFAHTDH 63


RESULT 8
US-08-863-028-30
; Sequence 30, Application US/08863028
; Patent No. 5853991
;  GENERAL INFORMATION:
;    APPLICANT:  WANG, Xun
;    APPLICANT:  DUVICK, Jonathan P.
;    APPLICANT:  BRIGGS, Steven P.
;    TITLE OF INVENTION:  PCR-BASED CDNA SUBTRACTIVE CLONING
;    TITLE OF INVENTION:  METHOD
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/863,028
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/858,767
```

```
      FILING DATE:  19-MAY-1997
      APPLICATION NUMBER:  US 08/481,687
      FILING DATE:  07-JUN-1995
    ATTORNEY/AGENT INFORMATION:
      NAME:  BENT, Stephen A.
      REGISTRATION NUMBER:  29,768
      REFERENCE/DOCKET NUMBER:  33229/325/PIHI
    TELECOMMUNICATION INFORMATION:
      TELEPHONE:  (202)672-5300
      TELEFAX:  (202)672-5399
      TELEX:  904136
  INFORMATION FOR SEQ ID NO:  30:
    SEQUENCE CHARACTERISTICS:
      LENGTH:  151 amino acids
      TYPE:  amino acid
      TOPOLOGY:  linear
    MOLECULE TYPE:  protein
'S-08-863-028-30

 Query Match             75.9%;  Score 22;  DB 2;  Length 151;
 Best Local Similarity   37.5%;  Pred. No. 8.9e+02;
 Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

'y          1 HXXXHXXH 8
              |   |  |
)b         56 HAFAHTDH 63


'ESULT 9
'S-09-615-192A-289
 Sequence 289, Application US/09615192A
 Patent No. 6410718
 GENERAL INFORMATION:
  APPLICANT:  Bloksberg, Leonard N.
  APPLICANT:  Havukkala, Ilkka
  TITLE OF INVENTION:  Materials and Methods for the
  TITLE OF INVENTION:  Modification of Plant Lignin Content
  FILE REFERENCE:  11000.1003c4U
  CURRENT APPLICATION NUMBER: US/09/615,192A
  CURRENT FILING DATE:  2000-07-12
  PRIOR APPLICATION NUMBER:  US 08/975,316
  PRIOR FILING DATE:  1997-11-21
  PRIOR APPLICATION NUMBER:  US 08/713,000
  PRIOR FILING DATE:  1996-09-11
  PRIOR APPLICATION NUMBER:  US 09/169,789
  PRIOR FILING DATE:  1998-10-09
  NUMBER OF SEQ ID NOS:  405
  SOFTWARE:  FastSEQ for Windows Version 3.0
 SEQ ID NO 289
   LENGTH: 179
   TYPE: PRT
   ORGANISM: Eucalyptus grandis
'S-09-615-192A-289

 Query Match             75.9%;  Score 22;  DB 4;  Length 179;
 Best Local Similarity   37.5%;  Pred. No. 1e+03;
 Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

)y          1 HXXXHXXH 8
              |   |  |
)b        170 HSIAHSDH 177


'ESULT 10
'S-09-252-991A-22610
 Sequence 22610, Application US/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:
  APPLICANT:  Marc J. Rubenfield et al.
  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
  FILE REFERENCE:  107196.136
  CURRENT APPLICATION NUMBER: US/09/252,991A
  CURRENT FILING DATE:  1999-02-18
  PRIOR APPLICATION NUMBER:  US 60/074,788
  PRIOR FILING DATE:  1998-02-18
  PRIOR APPLICATION NUMBER:  US 60/094,190
  PRIOR FILING DATE:  1998-07-27
  NUMBER OF SEQ ID NOS:  33142
 SEQ ID NO 22610
   LENGTH: 249
   TYPE: PRT
   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22610

 Query Match             75.9%;  Score 22;  DB 4;  Length 249;
 Best Local Similarity   37.5%;  Pred. No. 1.3e+03;
 Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db         60 HAAHHHAH 67


RESULT 11
US-09-252-991A-20987
; Sequence 20987, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 20987
;   LENGTH: 260
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (78)
;   OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-252-991A-20987

 Query Match             75.9%;  Score 22;  DB 4;  Length 260;
 Best Local Similarity   37.5%;  Pred. No. 1.3e+03;
 Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db        113 HLAAHRSH 120


RESULT 12
US-08-858-207A-447
; Sequence 447, Application US/08858207A
; Patent No. 6348328
;  GENERAL INFORMATION:
;    APPLICANT:  Black, Michael
;    APPLICANT:  Hodgson, John
;    APPLICANT:  Knowles, David
;    APPLICANT:  Nicholas, Richard
;    APPLICANT:  Stodola, Robert
;    TITLE OF INVENTION:  No. 6348328el Compounds
;    NUMBER OF SEQUENCES:  552
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
```

```
      COUNTRY:  USA
      ZIP:  19406-0939
    COMPUTER READABLE FORM:
      MEDIUM TYPE:  Diskette
      COMPUTER:  IBM Compatible
      OPERATING SYSTEM:  DOS
      SOFTWARE:  FastSEQ for Windows Version 2.0
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER:  US/08/858,207A
      FILING DATE:  09-MAY-1997
      CLASSIFICATION:  435
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  60/017670
      FILING DATE:  14-MAY-1996
    ATTORNEY/AGENT INFORMATION:
      NAME:  Gimmi, Edward R
      REGISTRATION NUMBER:  38,891
      REFERENCE/DOCKET NUMBER:  P50475
    TELECOMMUNICATION INFORMATION:
      TELEPHONE:  610-270-4478
      TELEFAX:  610-270-5090
      TELEX:
  INFORMATION FOR SEQ ID NO:  447:
    SEQUENCE CHARACTERISTICS:
      LENGTH:  272 amino acids
      TYPE:  amino acid
      STRANDEDNESS:  single
      TOPOLOGY:  linear
    MOLECULE TYPE:  No. 6348328e
'S-08-858-207A-447

 Query Match             75.9%;  Score 22;  DB 4;  Length 272;
 Best Local Similarity   37.5%;  Pred. No. 1.4e+03;
 Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

'y          1 HXXXHXXH 8
              |   |  |
'b        151 HTATHLLH 158


ESULT 13
'S-09-328-352-6181
 Sequence 6181, Application US/09328352
 Patent No. 6562958
 GENERAL INFORMATION:
  APPLICANT: Gary L. Breton et al.
  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
  FILE REFERENCE: GTC99-03PA
  CURRENT APPLICATION NUMBER: US/09/328,352
  CURRENT FILING DATE: 1999-06-04
  NUMBER OF SEQ ID NOS: 8252
 SEQ ID NO 6181
   LENGTH: 323
   TYPE: PRT
   ORGANISM: Acinetobacter baumannii
'S-09-328-352-6181

 Query Match             75.9%;  Score 22;  DB 4;  Length 323;
 Best Local Similarity   37.5%;  Pred. No. 1.6e+03;
 Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

'y          1 HXXXHXXH 8
              |   |  |
'b        303 HIAQHASH 310


:ESULT 14
'S-08-758-621-4
 Sequence 4, Application US/08758621
 Patent No. 5846821
  GENERAL INFORMATION:
    APPLICANT:  Guerinot, Mary Lou, and  Eide, David J.
;    TITLE OF INVENTION:  Metal-Regulated Transporters and Uses Therefor
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/758,621
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/018,578
;      FILING DATE:  29-MAY-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silveri, Jean M.
;      REGISTRATION NUMBER:  39,030
;      REFERENCE/DOCKET NUMBER:  DCI-099CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  355 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-758-621-4

  Query Match             75.9%;  Score 22;  DB 2;  Length 355;
  Best Local Similarity   37.5%;  Pred. No. 1.7e+03;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db        178 HIHTHASH 185


RESULT 15
US-09-107-858-4
; Sequence 4, Application US/09107858
; Patent No. 6162900
; GENERAL INFORMATION:
;  APPLICANT: Guerinot, Mary Lou et al.
;  TITLE OF INVENTION: METAL-REGULATED TRANSPORTERS AND USES THEREFOR
;  FILE REFERENCE: DCI-099CPDV
;  CURRENT APPLICATION NUMBER: US/09/107,858
;  CURRENT FILING DATE: 1998-06-30
;  EARLIER APPLICATION NUMBER: 08/758,621
;  EARLIER FILING DATE: 1996-11-27
;  NUMBER OF SEQ ID NOS: 27
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 355
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
US-09-107-858-4

  Query Match             75.9%;  Score 22;  DB 3;  Length 355;
  Best Local Similarity   37.5%;  Pred. No. 1.7e+03;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db        178 HIHTHASH 185
```

earch completed: November 21, 2003, 15:51:03
ob time : 14.5 secs

GenCore version 5.1.6


M protein - protein search, using sw model

Run on: November 21, 2003, 15:49:51 ; Search time 23.5 Seconds
(without alignments)
62.148 Million cell updates/sec

Title: US-10-064-903-1
Perfect score: 29
Sequence: 1 HXXXHXXH 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 23 | 79.3 | 890 | 15 | US-10-156-761-14378 | Sequence 14378, A |
| 2 | 22 | 75.9 | 61 | 14 | US-10-001-835-228 | Sequence 228, App |
| 3 | 22 | 75.9 | 61 | 15 | US-10-106-698-6639 | Sequence 6639, Ap |
| 4 | 22 | 75.9 | 69 | 9 | US-09-864-761-35891 | Sequence 35891, A |
| 5 | 22 | 75.9 | 69 | 12 | US-10-029-386-29728 | Sequence 29728, A |
| 6 | 22 | 75.9 | 73 | 12 | US-10-238-075-517 | Sequence 517, App |
| 7 | 22 | 75.9 | 102 | 12 | US-10-231-417-489 | Sequence 489, App |
| 8 | 22 | 75.9 | 104 | 10 | US-09-764-864-1330 | Sequence 1330, Ap |
| 9 | 22 | 75.9 | 110 | 9 | US-09-864-761-35339 | Sequence 35339, A |
| 10 | 22 | 75.9 | 110 | 9 | US-09-864-761-46752 | Sequence 46752, A |
| 11 | 22 | 75.9 | 114 | 9 | US-09-864-761-37988 | Sequence 37988, A |
| 12 | 22 | 75.9 | 136 | 10 | US-09-893-737-74 | Sequence 74, Appl |
| 13 | 22 | 75.9 | 161 | 15 | US-10-213-880-4 | Sequence 4, Appli |
| 14 | 22 | 75.9 | 167 | 9 | US-09-864-761-34765 | Sequence 34765, A |
| 15 | 22 | 75.9 | 179 | 16 | US-10-174-693-289 | Sequence 289, App |
| 16 | 22 | 75.9 | 180 | 9 | US-09-811-284-249 | Sequence 249, App |
| 17 | 22 | 75.9 | 193 | 9 | US-09-191-687B-4 | Sequence 4, Appli |
| 18 | 22 | 75.9 | 193 | 15 | US-10-228-796-4 | Sequence 4, Appli |
| 19 | 22 | 75.9 | 207 | 9 | US-09-804-551B-42 | Sequence 42, Appl |
| 20 | 22 | 75.9 | 221 | 12 | US-10-032-585-7060 | Sequence 7060, Ap |
| 21 | 22 | 75.9 | 340 | 9 | US-09-971-361-10 | Sequence 10, Appl |
| 22 | 22 | 75.9 | 342 | 15 | US-10-156-761-12399 | Sequence 12399, A |
| 23 | 22 | 75.9 | 352 | 15 | US-10-232-563-2 | Sequence 2, Appli |
| 24 | 22 | 75.9 | 359 | 15 | US-10-232-563-6 | Sequence 6, Appli |
| 25 | 22 | 75.9 | 359 | 15 | US-10-232-563-7 | Sequence 7, Appli |
| 26 | 22 | 75.9 | 385 | 12 | US-09-855-612-2 | Sequence 2, Appli |
| 27 | 22 | 75.9 | 385 | 14 | US-10-139-262-2 | Sequence 2, Appli |
| 28 | 22 | 75.9 | 385 | 15 | US-10-255-969-2 | Sequence 2, Appli |
| 29 | 22 | 75.9 | 397 | 10 | US-09-925-300-1531 | Sequence 1531, Ap |
| 30 | 22 | 75.9 | 408 | 9 | US-09-864-761-37954 | Sequence 37954, A |
| 31 | 22 | 75.9 | 420 | 12 | US-10-160-764-84 | Sequence 84, Appl |
| 32 | 22 | 75.9 | 424 | 15 | US-10-156-761-8087 | Sequence 8087, Ap |
| 33 | 22 | 75.9 | 483 | 10 | US-09-905-999-20 | Sequence 20, Appl |
| 34 | 22 | 75.9 | 500 | 12 | US-10-032-585-7530 | Sequence 7530, Ap |
| 35 | 22 | 75.9 | 527 | 15 | US-10-128-714-3378 | Sequence 3378, Ap |
| 36 | 22 | 75.9 | 556 | 15 | US-10-128-714-3561 | Sequence 3561, Ap |
| 37 | 22 | 75.9 | 556 | 15 | US-10-128-714-8378 | Sequence 8378, Ap |
| 38 | 22 | 75.9 | 556 | 15 | US-10-128-714-8561 | Sequence 8561, Ap |
| 39 | 22 | 75.9 | 559 | 9 | US-09-821-016-1 | Sequence 1, Appli |
| 40 | 22 | 75.9 | 559 | 9 | US-09-820-952A-1 | Sequence 1, Appli |
| 41 | 22 | 75.9 | 559 | 9 | US-09-820-721A-1 | Sequence 1, Appli |
| 42 | 22 | 75.9 | 559 | 10 | US-09-364-847-21 | Sequence 21, Appl |
| 43 | 22 | 75.9 | 559 | 15 | US-10-218-519-1 | Sequence 1, Appli |
| 44 | 22 | 75.9 | 559 | 15 | US-10-259-632-1 | Sequence 1, Appli |
| 45 | 22 | 75.9 | 559 | 15 | US-10-266-787-1 | Sequence 1, Appli |

ALIGNMENTS

RESULT 1
US-10-156-761-14378
; Sequence 14378, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 14378
; LENGTH: 890
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-14378

Query Match 79.3%; Score 23; DB 15; Length 890;
Best Local Similarity 37.5%; Pred. No. 6e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 HXXXHXXH 8
| | | |
Db 573 HSATHLTH 580

RESULT 2
US-10-001-835-228

Sequence 228, Application US/10001835
Publication No. US20020160387A1
GENERAL INFORMATION:
APPLICANT: Salceda, Susana
APPLICANT: Macina, Roberto
APPLICANT: Recipon, Herve
APPLICANT: Cafferkey, Robert
APPLICANT: Sun, Yongming
APPLICANT: Liu, Chenghua
TITLE OF INVENTION: Compositions and Methods Relating to Ovary Specific Genes and Prc
FILE REFERENCE: DEX-0277
CURRENT APPLICATION NUMBER: US/10/001,835
CURRENT FILING DATE: 2001-11-20
PRIOR APPLICATION NUMBER: 60/249,997
PRIOR FILING DATE: 2000-11-20
NUMBER OF SEQ ID NOS: 228
SOFTWARE: PatentIn version 3.1
SEQ ID NO 228
LENGTH: 61
TYPE: PRT
ORGANISM: Homo sapien
S-10-001-835-228

Query Match 75.9%; Score 22; DB 14; Length 61;
Best Local Similarity 37.5%; Pred. No. 1.3e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 HXXXHXXH 8
| | | |
Db 22 HRSTHQAH 29

ESULT 3
S-10-106-698-6639
Sequence 6639, Application US/10106698
Publication No. US20030109690A1
GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
FILE REFERENCE: PA005P1
CURRENT APPLICATION NUMBER: US/10/106,698
CURRENT FILING DATE: 2002-03-27
PRIOR APPLICATION NUMBER: PCT/US00/26524
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US 60/157,137
PRIOR FILING DATE: 1999-09-29
PRIOR APPLICATION NUMBER: US 60/163,280
PRIOR FILING DATE: 1999-11-03
NUMBER OF SEQ ID NOS: 8564
SOFTWARE: PatentIn Ver. 3.0
SEQ ID NO 6639
LENGTH: 61
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (24)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (28)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (49)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (53)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (61)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
S-10-106-698-6639

Query Match 75.9%; Score 22; DB 15; Length 61;
Best Local Similarity 37.5%; Pred. No. 1.3e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 HXXXHXXH 8
| | | |
Db 41 HASDHFAH 48

RESULT 4
US-09-864-761-35891
; Sequence 35891, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 35891
; LENGTH: 69
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC009743.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
; OTHER INFORMATION: EST_HUMAN HIT: AA641863.1, EVALUE 7.10e-01
; OTHER INFORMATION: SWISSPROT HIT: P04929, EVALUE 3.30e+00
US-09-864-761-35891

```
 Query Match             75.9%;  Score 22;  DB 9;  Length 69;
 Best Local Similarity   37.5%;  Pred. No. 1.4e+03;
 Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |  |  |
Db         51 HTVQHTSH 58


ESULT 5
S-10-029-386-29728
 Sequence 29728, Application US/10029386
 Publication No. US20030194704A1
 GENERAL INFORMATION:
  APPLICANT: Penn, Sharron G.
  APPLICANT:  Rank, David R.
  APPLICANT:  Hanzel, David K.
  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
  TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
  FILE REFERENCE: AEOMICA-X-2
  CURRENT APPLICATION NUMBER: US/10/029,386
  CURRENT FILING DATE:  2001-12-20
  NUMBER OF SEQ ID NOS: 34288
  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
 SEQ ID NO 29728
   LENGTH: 69
   TYPE: PRT
   ORGANISM: Homo sapiens
   FEATURE:
   OTHER INFORMATION: MAP TO CHR11.1
   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2
   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.6
   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.67
   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
   OTHER INFORMATION: SWISSPROT HIT: P22932, EVALUE 1.20e+00
S-10-029-386-29728

 Query Match             75.9%;  Score 22;  DB 12;  Length 69;
 Best Local Similarity   37.5%;  Pred. No. 1.4e+03;
 Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |  |  |
Db         52 HDAAHSGH 59


ESULT 6
S-10-238-075-517
 Sequence 517, Application US/10238075
 Publication No. US20030148324A1
 GENERAL INFORMATION:
  APPLICANT: I.N.S.E.R.M.
  TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolat
  TITLE OF INVENTION:  E.coli, and biological uses of these polynucleotides and of thei
  FILE REFERENCE: BLANDINE
  CURRENT APPLICATION NUMBER: US/10/238,075
  CURRENT FILING DATE:  2002-09-10
  PRIOR APPLICATION NUMBER: 0003145
  PRIOR FILING DATE: 2000-03-10
  NUMBER OF SEQ ID NOS: 1576
  SOFTWARE: PatentIn version 3.1
 SEQ ID NO 517
   LENGTH: 73
   TYPE: PRT
   ORGANISM: Escherichia coli
S-10-238-075-517

 Query Match             75.9%;  Score 22;  DB 12;  Length 73;
 Best Local Similarity   37.5%;  Pred. No. 1.5e+03;
 Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |  |  |
Db         15 HSHQHTAH 22


RESULT 7
US-10-231-417-489
; Sequence 489, Application US/10231417
; Publication No. US20030176681A1
; GENERAL INFORMATION:
;  APPLICANT: Feng et al.
;  TITLE OF INVENTION: 148 Human Secreted Proteins
;  FILE REFERENCE: PZ019P1
;  CURRENT APPLICATION NUMBER: US/10/231,417
;  CURRENT FILING DATE:  2002-08-30
;  PRIOR APPLICATION NUMBER: US/09/296,622
;  PRIOR FILING DATE: 1999-04-23
;  NUMBER OF SEQ ID NOS: 619
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 489
;   LENGTH: 102
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-231-417-489

  Query Match             75.9%;  Score 22;  DB 12;  Length 102;
  Best Local Similarity   37.5%;  Pred. No. 1.9e+03;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |  |  |
Db         56 HTHTHTGH 63


RESULT 8
US-09-764-864-1330
; Sequence 1330, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ23
;  CURRENT APPLICATION NUMBER: US/09/764,864
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1792
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1330
;   LENGTH: 104
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (65)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (77)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (102)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1330

  Query Match             75.9%;  Score 22;  DB 10;  Length 104;
  Best Local Similarity   37.5%;  Pred. No. 1.9e+03;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |  |  |
Db         81 HTRAHIAH 88


RESULT 9
US-09-864-761-35339
```

```
Sequence 35339, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
 APPLICANT: Penn, Sharron G.
 APPLICANT:  Rank, David R.
 APPLICANT:  Hanzel, David K.
 APPLICANT:  Chen, Wensheng
 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
 FILE REFERENCE: Aeomica-X-1
 CURRENT APPLICATION NUMBER: US/09/864,761
 CURRENT FILING DATE:  2001-05-23
 PRIOR APPLICATION NUMBER: US 60/180,312
 PRIOR FILING DATE: 2000-02-04
 PRIOR APPLICATION NUMBER: US 60/207,456
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: US 09/632,366
 PRIOR FILING DATE: 2000-08-03
 PRIOR APPLICATION NUMBER: GB 24263.6
 PRIOR FILING DATE: 2000-10-04
 PRIOR APPLICATION NUMBER: US 60/236,359
 PRIOR FILING DATE: 2000-09-27
 PRIOR APPLICATION NUMBER: PCT/US01/00666
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00667
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00664
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00669
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00665
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00668
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00663
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00662
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00661
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00670
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: US 60/234,687
 PRIOR FILING DATE: 2000-09-21
 PRIOR APPLICATION NUMBER: US 09/608,408
 PRIOR FILING DATE: 2000-06-30
 PRIOR APPLICATION NUMBER: US 09/774,203
 PRIOR FILING DATE: 2001-01-29
 NUMBER OF SEQ ID NOS: 49117
 SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 35339
  LENGTH: 110
  TYPE: PRT
  ORGANISM: Homo sapiens
  FEATURE:
  OTHER INFORMATION: MAP TO AP000507.1
  OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.4
  OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.7
  OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.4
  OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 6.3
  OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.6
  OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.6
  OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.5
  OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.3
  OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =   4
  OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
  OTHER INFORMATION: SWISSPROT HIT: Q30201, EVALUE 9.00e-16
  OTHER INFORMATION: EST_HUMAN HIT: BE877225.1, EVALUE 8.00e-54
US-09-864-761-35339

 Query Match           75.9%;  Score 22;  DB 9;  Length 110;
 Best Local Similarity  37.5%;  Pred. No. 2e+03;
 Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |  |  |
Db         90 HSGNHSTH 97


RESULT 10
US-09-864-761-46752
; Sequence 46752, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 46752
;   LENGTH: 110
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC007389.2
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.5
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.84
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL =   1
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
;   OTHER INFORMATION: EST_HUMAN HIT: AA305279.1, EVALUE 4.00e-23
```

```
   OTHER INFORMATION: SWISSPROT HIT: Q26609, EVALUE 8.00e+00
S-09-864-761-46752

 Query Match             75.9%;  Score 22;  DB 9;  Length 110;
 Best Local Similarity   37.5%;  Pred. No. 2e+03;
 Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db         55 HQATHSRH 62


ESULT 11
S-09-864-761-37988
 Sequence 37988, Application US/09864761
 Patent No. US20020048763A1
 GENERAL INFORMATION:
  APPLICANT: Penn, Sharron G.
  APPLICANT:  Rank, David R.
  APPLICANT:  Hanzel, David K.
  APPLICANT:  Chen, Wensheng
  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
  FILE REFERENCE: Aeomica-X-1
  CURRENT APPLICATION NUMBER: US/09/864,761
  CURRENT FILING DATE:  2001-05-23
  PRIOR APPLICATION NUMBER: US 60/180,312
  PRIOR FILING DATE: 2000-02-04
  PRIOR APPLICATION NUMBER: US 60/207,456
  PRIOR FILING DATE: 2000-05-26
  PRIOR APPLICATION NUMBER: US 09/632,366
  PRIOR FILING DATE: 2000-08-03
  PRIOR APPLICATION NUMBER: GB 24263.6
  PRIOR FILING DATE: 2000-10-04
  PRIOR APPLICATION NUMBER: US 60/236,359
  PRIOR FILING DATE: 2000-09-27
  PRIOR APPLICATION NUMBER: PCT/US01/00666
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00667
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00664
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00669
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00665
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00668
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00663
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00662
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00661
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00670
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: US 60/234,687
  PRIOR FILING DATE: 2000-09-21
  PRIOR APPLICATION NUMBER: US 09/608,408
  PRIOR FILING DATE: 2000-06-30
  PRIOR APPLICATION NUMBER: US 09/774,203
  PRIOR FILING DATE: 2001-01-29
  NUMBER OF SEQ ID NOS: 49117
  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
 SEQ ID NO 37988
   LENGTH: 114
   TYPE: PRT
   ORGANISM: Homo sapiens
   FEATURE:
   OTHER INFORMATION: MAP TO D84394.1
   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6
   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.9
;  OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.3
;  OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.3
;  OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
;  OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.4
;  OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4
;  OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.99
;  OTHER INFORMATION: EST_HUMAN HIT: BE743982.1, EVALUE 8.00e-54
;  OTHER INFORMATION: SWISSPROT HIT: Q30201, EVALUE 9.00e-16
US-09-864-761-37988

  Query Match             75.9%;  Score 22;  DB 9;  Length 114;
  Best Local Similarity   37.5%;  Pred. No. 2e+03;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db         94 HSGNHSTH 101


RESULT 12
US-09-893-737-74
; Sequence 74, Application US/09893737
; Patent No. US20020110855A1
; GENERAL INFORMATION:
;  APPLICANT: Sheppard, Paul O.
;  APPLICANT:  Presnell, Scott R.
;  TITLE OF INVENTION: MAMMALIAN SECRETED PROTEINS
;  FILE REFERENCE: 00-41
;  CURRENT APPLICATION NUMBER: US/09/893,737
;  CURRENT FILING DATE:  2001-06-28
;  PRIOR APPLICATION NUMBER: US 60/215,446
;  PRIOR FILING DATE: 2000-06-30
;  NUMBER OF SEQ ID NOS: 329
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 74
;   LENGTH: 136
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-893-737-74

  Query Match             75.9%;  Score 22;  DB 10;  Length 136;
  Best Local Similarity   37.5%;  Pred. No. 2.3e+03;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db        109 HGTAHARH 116


RESULT 13
US-10-213-880-4
; Sequence 4, Application US/10213880
; Publication No. US20030088083A1
; GENERAL INFORMATION:
;  APPLICANT: Allen, Stephen M.
;  APPLICANT:  Famodu, Omolayo O.
;  APPLICANT:  Rasco-Gaunt, Sonriza
;  APPLICANT:  Thorpe, Catherine J.
;  TITLE OF INVENTION: Metal-Binding Proteins
;  FILE REFERENCE: BB1513 US NA
;  CURRENT APPLICATION NUMBER: US/10/213,880
;  CURRENT FILING DATE:  2002-08-07
;  PRIOR APPLICATION NUMBER: U.S. 60/310,522
;  PRIOR FILING DATE: 2001-08-07
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
;   LENGTH: 161
;   TYPE: PRT
;   ORGANISM: Momordica charantia
US-10-213-880-4

  Query Match             75.9%;  Score 22;  DB 15;  Length 161;
```

```
 Best Local Similarity   37.5%;  Pred. No. 2.6e+03;
 Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db         15 HSHSHSGH 22


RESULT 14
US-09-864-761-34765
 Sequence 34765, Application US/09864761
 Patent No. US20020048763A1
 GENERAL INFORMATION:
  APPLICANT: Penn, Sharron G.
  APPLICANT:  Rank, David R.
  APPLICANT:  Hanzel, David K.
  APPLICANT:  Chen, Wensheng
  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
  FILE REFERENCE: Aeomica-X-1
  CURRENT APPLICATION NUMBER: US/09/864,761
  CURRENT FILING DATE:  2001-05-23
  PRIOR APPLICATION NUMBER: US 60/180,312
  PRIOR FILING DATE: 2000-02-04
  PRIOR APPLICATION NUMBER: US 60/207,456
  PRIOR FILING DATE: 2000-05-26
  PRIOR APPLICATION NUMBER: US 09/632,366
  PRIOR FILING DATE: 2000-08-03
  PRIOR APPLICATION NUMBER: GB 24263.6
  PRIOR FILING DATE: 2000-10-04
  PRIOR APPLICATION NUMBER: US 60/236,359
  PRIOR FILING DATE: 2000-09-27
  PRIOR APPLICATION NUMBER: PCT/US01/00666
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00667
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00664
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00669
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00665
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00668
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00663
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00662
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00661
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00670
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: US 60/234,687
  PRIOR FILING DATE: 2000-09-21
  PRIOR APPLICATION NUMBER: US 09/608,408
  PRIOR FILING DATE: 2000-06-30
  PRIOR APPLICATION NUMBER: US 09/774,203
  PRIOR FILING DATE: 2001-01-29
  NUMBER OF SEQ ID NOS: 49117
  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34765
;   LENGTH: 167
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC006371.2
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.6
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.5
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3.4
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.3
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.5
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL =   3
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.3
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.4
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.9
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.9
;   OTHER INFORMATION: EST_HUMAN HIT: BE958003.1, EVALUE 4.60e+00
US-09-864-761-34765

  Query Match             75.9%;  Score 22;  DB 9;  Length 167;
  Best Local Similarity   37.5%;  Pred. No. 2.7e+03;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db        112 HMHTHTSH 119


RESULT 15
US-10-174-693-289
; Sequence 289, Application US/10174693
; Publication No. US20030131373A1
; GENERAL INFORMATION:
;  APPLICANT: Bloksberg, Leonard N.
;  APPLICANT:  Havukkala, Ilkka
;  TITLE OF INVENTION: Materials and Methods for the
;  TITLE OF INVENTION:  Modification of Plant Lignin Content
;  FILE REFERENCE: 11000.1003c5
;  CURRENT APPLICATION NUMBER: US/10/174,693
;  CURRENT FILING DATE:  2002-06-18
;  PRIOR APPLICATION NUMBER: US 08/975,316
;  PRIOR FILING DATE: 1997-11-21
;  PRIOR APPLICATION NUMBER: US 08/713,000
;  PRIOR FILING DATE: 1996-09-11
;  PRIOR APPLICATION NUMBER: US 09/169,789
;  PRIOR FILING DATE: 1998-10-09
;  PRIOR APPLICATION NUMBER: US 09/615,192
;  PRIOR FILING DATE: 2000-07-12
;  NUMBER OF SEQ ID NOS: 407
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 289
;   LENGTH: 179
;   TYPE: PRT
;   ORGANISM: Eucalyptus grandis
US-10-174-693-289

  Query Match             75.9%;  Score 22;  DB 16;  Length 179;
  Best Local Similarity   37.5%;  Pred. No. 2.8e+03;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db        170 HSIAHSDH 177


Search completed: November 21, 2003, 15:58:28
Job time : 24.5 secs
```

GenCore version 5.1.6


M protein - protein search, using sw model

un on: November 21, 2003, 15:44:40 ; Search time 13 Seconds
(without alignments)
59.181 Million cell updates/sec

itle: US-10-064-903-1
erfect score: 29
equence: 1 HXXXHXXH 8

coring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

earched: 283308 seqs, 96168682 residues

otal number of hits satisfying chosen parameters: 283308

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| esult No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 23 | 79.3 | 152 | 2 | C72662 | hypothetical prote |
| 2 | 23 | 79.3 | 177 | 2 | T26468 | hypothetical prote |
| 3 | 23 | 79.3 | 240 | 2 | F82790 | GMP synthase XF056 |
| 4 | 23 | 79.3 | 327 | 2 | AC2120 | cytochrome c oxida |
| 5 | 23 | 79.3 | 342 | 2 | T15850 | hypothetical prote |
| 6 | 23 | 79.3 | 382 | 2 | T35709 | hypothetical prote |
| 7 | 23 | 79.3 | 472 | 2 | T27755 | hypothetical prote |
| 8 | 23 | 79.3 | 508 | 2 | S59870 | fork head domain p |
| 9 | 23 | 79.3 | 510 | 2 | S55124 | probable membrane |
| 10 | 23 | 79.3 | 606 | 2 | B69805 | conserved hypothet |
| 11 | 23 | 79.3 | 826 | 2 | T46060 | hypothetical prote |
| 12 | 23 | 79.3 | 826 | 2 | T46061 | hypothetical prote |
| 13 | 22 | 75.9 | 52 | 2 | S63324 | hypothetical prote |
| 14 | 22 | 75.9 | 61 | 2 | AC0287 | hypothetical prote |
| 15 | 22 | 75.9 | 121 | 2 | D82711 | hypothetical prote |
| 16 | 22 | 75.9 | 135 | 2 | I49275 | protein kinase STY |
| 17 | 22 | 75.9 | 144 | 2 | H75636 | transposase-relate |
| 18 | 22 | 75.9 | 177 | 2 | S65780 | glycine/proline-ri |
| 19 | 22 | 75.9 | 198 | 2 | B83717 | NADP-dependent alc |
| 20 | 22 | 75.9 | 208 | 2 | T35454 | hypothetical prote |
| 21 | 22 | 75.9 | 237 | 2 | S19103 | hypothetical prote |
| 22 | 22 | 75.9 | 263 | 2 | G75590 | hypothetical prote |
| 23 | 22 | 75.9 | 306 | 2 | I49068 | protein kinase STY |
| 24 | 22 | 75.9 | 312 | 2 | T27004 | hypothetical prote |
| 25 | 22 | 75.9 | 316 | 2 | D71375 | probable ABC trans |
| 26 | 22 | 75.9 | 325 | 2 | T44782 | exbB protein [impo |
| 27 | 22 | 75.9 | 337 | 1 | A42654 | alcohol dehydrogen |
| 28 | 22 | 75.9 | 339 | 1 | S45605 | alcohol dehydrogen |
| 29 | 22 | 75.9 | 339 | 1 | S47643 | alcohol dehydrogen |
| 30 | 22 | 75.9 | 340 | 2 | T37030 | alcohol dehydrogen |
| 31 | 22 | 75.9 | 341 | 2 | E83340 | hypothetical prote |
| 32 | 22 | 75.9 | 345 | 2 | T16935 | hypothetical prote |
| 33 | 22 | 75.9 | 355 | 2 | T52183 | zinc transporter Z |
| 34 | 22 | 75.9 | 364 | 2 | JC5800 | peptidylglycine mo |
| 35 | 22 | 75.9 | 383 | 2 | A55739 | (MIC) protein MHC |
| 36 | 22 | 75.9 | 384 | 2 | T23604 | hypothetical prote |
| 37 | 22 | 75.9 | 394 | 2 | E87606 | hypothetical prote |
| 38 | 22 | 75.9 | 403 | 2 | C96757 | hypothetical prote |
| 39 | 22 | 75.9 | 416 | 2 | A32947 | filaggrin precurso |
| 40 | 22 | 75.9 | 419 | 2 | JQ2254 | farnesyl-diphospha |
| 41 | 22 | 75.9 | 420 | 2 | G95107 | gamma-glutamyl pho |
| 42 | 22 | 75.9 | 420 | 2 | A97976 | glutamate-5-semial |
| 43 | 22 | 75.9 | 424 | 2 | T01383 | GTPase-activating |
| 44 | 22 | 75.9 | 427 | 2 | I51580 | XFKH2 protein - Af |
| 45 | 22 | 75.9 | 440 | 2 | B71293 | hypothetical prote |

ALIGNMENTS

RESULT 1
C72662
hypothetical protein APE0723 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C;Accession: C72662
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takaha
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Ku
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyru
A;Reference number: A72450; MUID:99310339; PMID:10382966
A;Accession: C72662
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-152 <KAW>
A;Cross-references: DDBJ:AP000060; NID:g5104188; PIDN:BAA79699.1; PID:d1043485; PID:g510
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE0723

Query Match 79.3%; Score 23; DB 2; Length 152;
Best Local Similarity 37.5%; Pred. No. 3e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | |
Db 15 HSTIHAAH 22

RESULT 2
T26468
hypothetical protein Y11D7A.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T26468
R;Steward, C.
submitted to the EMBL Data Library, October 1998
A;Reference number: Z20218
A;Accession: T26468
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-177 <WIL>
A;Cross-references: EMBL:AL032632; PIDN:CAA21589.1; GSPDB:GN00022; CESP:Y11D7A.1
A;Experimental source: clone Y11D7A
C;Genetics:
A;Gene: CESP:Y11D7A.1
A;Map position: 4
A;Introns: 48/1; 102/1; 128/1

Query Match 79.3%; Score 23; DB 2; Length 177;
Best Local Similarity 37.5%; Pred. No. 3.4e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy          1 HXXXHXXH 8
              |   |  |
Db        140 HTTVHSSH 147
```

RESULT 3
F82790
IMP synthase XF0560 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: F82790
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: F82790
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-240 <SIM>
A;Cross-references: GB:AE003903; GB:AE003849; NID:g9105416; PIDN:AAF83370.1; GSPDB:GN001
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
Submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF0560

  Query Match             79.3%;  Score 23;  DB 2;  Length 240;
  Best Local Similarity   37.5%;  Pred. No. 4.5e+02;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

```
Qy          1 HXXXHXXH 8
              |   |  |
Db        142 HFSAHATH 149
```

RESULT 4
AC2120
cytochrome c oxidase chain II [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AC2120
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AC2120
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-327 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB74213.1; PID:g17131606; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: coxB
C;Superfamily: cytochrome-c oxidase chain II; cytochrome-c oxidase chain II homology
C;Keywords: copper; electron transfer; membrane-associated complex; respiratory chain
F;214,249,253,260/Binding site: copper 1 (His, Cys, Cys, Met) #status predicted
F;249,251,253,257/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted
F;251/Binding site: magnesium (Glu) (shared with chain I) #status predicted

  Query Match             79.3%;  Score 23;  DB 2;  Length 327;
  Best Local Similarity   37.5%;  Pred. No. 6e+02;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

```
Qy          1 HXXXHXXH 8
              |   |  |
Db        126 HASAHVAH 133
```

RESULT 5
T15850
hypothetical protein C56C10.10 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T15850
R;Fulton, L.
submitted to the EMBL Data Library, June 1995
A;Description: The sequence of C. elegans cosmid C56C10.
A;Reference number: Z18417
A;Accession: T15850
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-342 <FUL>
A;Cross-references: EMBL:U29488; NID:g868238; PID:g868248; PIDN:AAA68778.1; CESP:C56C10.1
A;Experimental source: strain Bristol N2
C;Genetics:
A;Gene: CESP:C56C10.10
A;Introns: 51/2; 144/2; 204/3; 241/3; 295/3

  Query Match             79.3%;  Score 23;  DB 2;  Length 342;
  Best Local Similarity   37.5%;  Pred. No. 6.2e+02;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

```
Qy          1 HXXXHXXH 8
              |   |  |
Db        128 HSHAHTTH 135
```

RESULT 6
T35709
hypothetical protein SC7H1.14 SC7H1.14 - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C;Accession: T35709
R;Murphy, L.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, January 1998
A;Reference number: Z21548
A;Accession: T35709
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-382 <MUR>
A;Cross-references: EMBL:AL021411; PIDN:CAA16201.1; GSPDB:GN00070; SCOEDB:SC7H1.14
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SC7H1.14

  Query Match             79.3%;  Score 23;  DB 2;  Length 382;
  Best Local Similarity   37.5%;  Pred. No. 6.9e+02;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

```
Qy          1 HXXXHXXH 8
              |   |  |
Db        370 HAARHAAH 377
```

RESULT 7
T27755
hypothetical protein ZK1320.9 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 23-Dec-2002
C;Accession: T27755

;Berks, M.
ubmitted to the EMBL Data Library, December 1994
;Reference number: Z20414
;Accession: T27755
;Status: preliminary; translated from GB/EMBL/DDBJ
;Molecule type: DNA
;Residues: 1-472 <WIL>
;Cross-references: EMBL:Z46934; PIDN:CAA87047.1; GSPDB:GN00020; CESP:ZK1320.9
;Experimental source: clone ZK1320
;Genetics:
;Gene: CESP:ZK1320.9
;Map position: 2
;Introns: 19/2; 55/1; 106/1; 160/1; 186/1; 323/1; 411/3
;Superfamily: acetyl-CoA hydrolase

Query Match 79.3%; Score 23; DB 2; Length 472;
Best Local Similarity 37.5%; Pred. No. 8.3e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

y 1 HXXXHXXH 8
b 198 HTTVHSSH 205

ESULT 8
59870
ork head domain protein crocodile - fruit fly (Drosophila melanogaster)
;Species: Drosophila melanogaster
;Date: 19-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 21-Jul-2000
;Accession: S59870; A46178
;Haecker, U.; Kaufmann, E.; Hartmann, C.; Juergens, G.; Knoechel, W.; Jaeckle, H.
MBO J. 14, 5306-5317, 1995
;Title: The Drosophila fork head domain protein crocodile is required for the establish
;Reference number: S59870; MUID:96080166; PMID:7489720
;Accession: S59870
;Status: not compared with conceptual translation
;Molecule type: mRNA
;Residues: 1-508 <HAE>
;Hacker, U.; Grossniklaus, U.; Gehring, W.J.; Jackle, H.
roc. Natl. Acad. Sci. U.S.A. 89, 8754-8758, 1992
;Title: Developmentally regulated Drosophila gene family encoding the fork head domain.
;Reference number: A46178; MUID:92409595; PMID:1356269
;Accession: A46178
;Status: preliminary; not compared with conceptual translation
;Molecule type: nucleic acid
;Residues: 55-182 <HAC>
;Cross-references: GB:M96440; NID:g157425; PIDN:AAF02177.1; PID:g6042185
;Note: sequence extracted from NCBI backbone (NCBIP:114222)
;Genetics:
;Gene: croc
;Cross-references: FlyBase:FBgn0014143
;Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology
;70-161/Domain: fork head DNA-binding domain homology <FHD>

Query Match 79.3%; Score 23; DB 2; Length 508;
Best Local Similarity 37.5%; Pred. No. 8.9e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

y 1 HXXXHXXH 8
b 207 HMAAHAAH 214

ESULT 9
55124
robable membrane protein YMR177w - yeast (Saccharomyces cerevisiae)
;Alternate names: hypothetical protein YM8010.07
;Species: Saccharomyces cerevisiae
;Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 06-Feb-1998
;Accession: S55124
;Churcher, C.M.
ubmitted to the EMBL Data Library, June 1995
;Reference number: S55118
A;Accession: S55124
A;Molecule type: DNA
A;Residues: 1-510 <CHU>
A;Cross-references: EMBL:Z49808; NID:g854440; PID:g854447; MIPS:YMR177w
A;Experimental source: strain AB972
C;Genetics:
A;Gene: SGD:MMT1
A;Cross-references: SGD:S0004789; MIPS:YMR177w
A;Map position: 13R
C;Keywords: transmembrane protein
F;168-184/Domain: transmembrane #status predicted <TM1>
F;232-248/Domain: transmembrane #status predicted <TM2>
F;332-348/Domain: transmembrane #status predicted <TM3>
F;350-366/Domain: transmembrane #status predicted <TM4>

Query Match 79.3%; Score 23; DB 2; Length 510;
Best Local Similarity 37.5%; Pred. No. 8.9e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 HXXXHXXH 8
Db 137 HTHSHAAH 144

RESULT 10
B69805
conserved hypothetical protein yfiX - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C;Accession: B69805
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berterc
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallerc
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.;
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauel,
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle,
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror,
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: B69805
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-606 <KUN>
A;Cross-references: GB:Z99108; GB:AL009126; NID:g2633055; PIDN:CAB12672.1; PID:e1182833;
A;Experimental source: strain 168
C;Genetics:
A;Gene: yfiX

Query Match 79.3%; Score 23; DB 2; Length 606;
Best Local Similarity 37.5%; Pred. No. 1e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 HXXXHXXH 8
Db 244 HSTSHITH 251

RESULT 11
T46060
hypothetical protein T18N14.20 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C;Accession: T46060
R;Delseny, M.; Berger, C.; Cooke, R.; Grellet, F.; Laudie, M.; Mewes, H.W.; Lemcke, K.; M
submitted to the Protein Sequence Database, December 1999
A;Reference number: Z23013

;Accession: T46060
;Status: preliminary
;Molecule type: DNA
;Residues: 1-826 <DEL>
;Cross-references: EMBL:AL132968
;Experimental source: cultivar Columbia; BAC clone T18N14
;Genetics:
;Map position: 3
;Introns: 476/3; 796/2
;Note: T18N14.20

Query Match 79.3%; Score 23; DB 2; Length 826;
Best Local Similarity 37.5%; Pred. No. 1.4e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

y 1 HXXXHXXH 8
| | |
b 446 HTYAHSSH 453

ESULT 12
46061
ypothetical protein T18N14.30 - Arabidopsis thaliana
;Species: Arabidopsis thaliana (mouse-ear cress)
;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
;Accession: T46061
;Delseny, M.; Berger, C.; Cooke, R.; Grellet, F.; Laudie, M.; Mewes, H.W.; Lemcke, K.;
ubmitted to the Protein Sequence Database, December 1999
.;Reference number: Z23013
.;Accession: T46061
.;Status: preliminary
.;Molecule type: DNA
.;Residues: 1-826 <DEL>
.;Cross-references: EMBL:AL132968
.;Experimental source: cultivar Columbia; BAC clone T18N14
';Genetics:
.;Map position: 3
.;Introns: 476/3; 796/2
.;Note: T18N14.30

Query Match 79.3%; Score 23; DB 2; Length 826;
Best Local Similarity 37.5%; Pred. No. 1.4e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

)y 1 HXXXHXXH 8
| | |
)b 446 HTYAHSSH 453

:ESULT 13
:63324
ypothetical protein YNL338w - yeast (Saccharomyces cerevisiae)
[;Alternate names: hypothetical protein N0170
:;Species: Saccharomyces cerevisiae
';Date: 27-Apr-1996 #sequence_revision 03-May-1996 #text_change 19-Apr-2002
:;Accession: S63324
:;Obermaier, B.; Piravandi, E.; Rinke, M.
ubmitted to the Protein Sequence Database, April 1996
\;Reference number: S63317
\;Accession: S63324
\;Molecule type: DNA
\;Residues: 1-52 <OBE>
\;Cross-references: EMBL:Z71614; NID:g1302466; PID:e239576; PID:g1302467; GSPDB:GN00014;
\;Experimental source: strain S288C
:;Genetics:
\;Gene: MIPS:YNL338w
\;Cross-references: SGD:S0005282
\;Map position: 14L

Query Match 75.9%; Score 22; DB 2; Length 52;
Best Local Similarity 37.5%; Pred. No. 1.7e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 HXXXHXXH 8
| | |
Db 38 HTHTHTHH 45

RESULT 14
AC0287
hypothetical protein YPO2354 [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C;Accession: AC0287
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; E
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, E
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AC0287
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-61 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC91159.1; PID:g15980351; GSPDB:GN00175
C;Genetics:
A;Gene: YPO2354

Query Match 75.9%; Score 22; DB 2; Length 61;
Best Local Similarity 37.5%; Pred. No. 2e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 HXXXHXXH 8
| | |
Db 47 HITHHTSH 54

RESULT 15
D82711
hypothetical protein XF1205 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: D82711
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequenc
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: D82711
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-121 <SIM>
A;Cross-references: GB:AE003954; GB:AE003849; NID:g9106165; PIDN:AAF84015.1; GSPDB:GN001
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Al
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H.
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigre
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaki
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Za
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF1205

Query Match 75.9%; Score 22; DB 2; Length 121;
Best Local Similarity 37.5%; Pred. No. 3.7e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 HXXXHXXH 8

```
                |   |   |
)b          63 HTFTHTEH 70

)earch completed: November 21, 2003, 15:50:23
Job time : 15 secs
```

GenCore version 5.1.6


M protein - protein search, using sw model

un on: November 21, 2003, 15:30:05 ; Search time 10 Seconds
(without alignments)
37.621 Million cell updates/sec

itle: US-10-064-903-1
erfect score: 29
equence: 1 HXXXHXXH 8

coring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

earched: 127863 seqs, 47026705 residues

otal number of hits satisfying chosen parameters: 127863

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| esult No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 23 | 79.3 | 508 | 1 | CROC_DROME | P32027 drosophila |
| 2 | 23 | 79.3 | 510 | 1 | YM43_YEAST | Q03218 saccharomyc |
| 3 | 23 | 79.3 | 880 | 1 | BRC4_DROME | Q24206 drosophila |
| 4 | 23 | 79.3 | 890 | 1 | SYA_STRCO | Q9kxp9 streptomyce |
| 5 | 23 | 79.3 | 1509 | 1 | GSR1_HUMAN | Q9nzm4 homo sapien |
| 6 | 22 | 75.9 | 52 | 1 | YN78_YEAST | P53820 saccharomyc |
| 7 | 22 | 75.9 | 316 | 1 | Y034_TREPA | O83077 treponema p |
| 8 | 22 | 75.9 | 337 | 1 | ADH1_BACST | P12311 bacillus st |
| 9 | 22 | 75.9 | 339 | 1 | ADH2_BACST | P42327 bacillus st |
| 10 | 22 | 75.9 | 339 | 1 | ADH3_BACST | P42328 bacillus st |
| 11 | 22 | 75.9 | 416 | 1 | FILA_HUMAN | P20930 homo sapien |
| 12 | 22 | 75.9 | 419 | 1 | PFTB_PEA | Q04903 pisum sativ |
| 13 | 22 | 75.9 | 420 | 1 | PROA_NEIMB | Q9jzg3 neisseria m |
| 14 | 22 | 75.9 | 420 | 1 | PROA_STRPN | Q97r94 streptococc |
| 15 | 22 | 75.9 | 427 | 1 | FKH2_XENLA | P32315 xenopus lae |
| 16 | 22 | 75.9 | 440 | 1 | Y693_TREPA | O83691 treponema p |
| 17 | 22 | 75.9 | 472 | 1 | SX14_DROME | P40656 drosophila |
| 18 | 22 | 75.9 | 483 | 1 | CLK1_MOUSE | P22518 mus musculu |
| 19 | 22 | 75.9 | 539 | 1 | DOP2_DROME | Q24563 drosophila |
| 20 | 22 | 75.9 | 559 | 1 | PHAA_PSEOL | P26494 pseudomonas |
| 21 | 22 | 75.9 | 590 | 1 | SYT3_HUMAN | Q9bqg1 homo sapien |
| 22 | 22 | 75.9 | 594 | 1 | SYA_BORBU | O51238 borrelia bu |
| 23 | 22 | 75.9 | 596 | 1 | FRDA_SHEFR | Q02469 shewanella |
| 24 | 22 | 75.9 | 605 | 1 | SYA_TREPA | O83980 treponema p |
| 25 | 22 | 75.9 | 679 | 1 | TKT1_YEAST | P23254 saccharomyc |
| 26 | 22 | 75.9 | 787 | 1 | AGL2_BACTQ | Q9f234 bacillus th |
| 27 | 22 | 75.9 | 842 | 1 | SYA_CAMJE | Q9pi05 campylobact |
| 28 | 22 | 75.9 | 860 | 1 | SYA_VIBCH | Q56648 vibrio chol |
| 29 | 22 | 75.9 | 860 | 1 | SYA_VIBPA | Q87lr3 vibrio para |
| 30 | 22 | 75.9 | 860 | 1 | SYA_VIBVU | Q8dc49 vibrio vuln |
| 31 | 22 | 75.9 | 863 | 1 | SYA_THEMA | Q9x1b6 thermotoga |
| 32 | 22 | 75.9 | 867 | 1 | SYA_AQUAE | O67323 aquifex aeo |
| 33 | 22 | 75.9 | 867 | 1 | SYA_FUSNN | Q8rfj8 fusobacteri |
| 34 | 22 | 75.9 | 871 | 1 | SYA_AQUPY | Q9xdm3 aquifex pyr |
| 35 | 22 | 75.9 | 872 | 1 | SYA_LACLA | Q9cew0 lactococcus |
| 36 | 22 | 75.9 | 872 | 1 | SYA_STRP8 | Q8p0e6 streptococc |
| 37 | 22 | 75.9 | 872 | 1 | SYA_STRPN | Q97q48 streptococc |
| 38 | 22 | 75.9 | 872 | 1 | SYA_STRPY | Q99z57 streptococc |
| 39 | 22 | 75.9 | 873 | 1 | SYA_WIGBR | Q8d2w8 wiggleswort |
| 40 | 22 | 75.9 | 874 | 1 | SYA_HAEIN | P43815 haemophilus |
| 41 | 22 | 75.9 | 874 | 1 | SYA_NEIMA | Q9jtg4 neisseria m |
| 42 | 22 | 75.9 | 874 | 1 | SYA_NEIMB | Q9jyg6 neisseria m |
| 43 | 22 | 75.9 | 874 | 1 | SYA_PASMU | P57933 pasteurella |
| 44 | 22 | 75.9 | 874 | 1 | SYA_PSEAE | Q9i553 pseudomonas |
| 45 | 22 | 75.9 | 875 | 1 | SYA_YERPE | Q8zbt8 yersinia pe |

ALIGNMENTS

RESULT 1
CROC_DROME
ID CROC_DROME STANDARD; PRT; 508 AA.
AC P32027; Q9VP32;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Fork head domain protein crocodile (FKH protein FD1).
GN CROC OR FD78E OR FD1 OR CG5069.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S;
RX MEDLINE=96080166; PubMed=7489720;
RA Haecker U., Kaufmann E., Hartmann C., Juergens G., Knoechel W.,
RA Jaeckle H.;
RT "The Drosophila fork head domain protein crocodile is required for
RT the establishment of head structures.";
RL EMBO J. 14:5306-5317(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

```
A   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
A   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
A   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
A   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
A   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
A   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
A   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
T   "The genome sequence of Drosophila melanogaster.";
L   Science 287:2185-2195(2000).
N   [3]
P   SEQUENCE OF 55-182 FROM N.A., FUNCTION, TISSUE SPECIFICITY, AND
P   DEVELOPMENTAL STAGE.
X   MEDLINE=92409595; PubMed=1356269;
A   Haecker U., Grossniklaus U., Gehring W.J., Jaeckle H.;
T   "Developmentally regulated Drosophila gene family encoding the fork
T   head domain.";
L   Proc. Natl. Acad. Sci. U.S.A. 89:8754-8758(1992).
C   -!- FUNCTION: REQUIRED FOR THE ESTABLISHMENT OF HEAD STRUCTURES.
C       REQUIRED TO FUNCTION AS AN EARLY PATTERNING GENE IN THE ANTERIOR-
C       MOST BLASTODERM HEAD SEGMENT ANLAGE AND FOR THE ESTABLISHMENT OF A
C       SPECIFIC HEAD SKELETAL STRUCTURE THAT DERIVES FROM THE NON-
C       ADJACENT INTERCALARY SEGMENT AT A LATER STAGE OF EMBRYOGENESIS.
C       BINDS THE CONSENSUS DNA SEQUENCE 5'-[AG]TAAA[TC]A-3'.
C   -!- SUBCELLULAR LOCATION: Nuclear.
C   -!- TISSUE SPECIFICITY: EXPRESSED IN EARLY BLASTODERM EMBRYOS IN
C       ANTERIOR AND POSTERIOR GUT PRECURSORS, AND, LATER IN A SUBSET OF
C       CELLS IN CENTRAL NERVOUS SYSTEM.
C   -!- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT EMBRYOGENESIS, MAXIMALLY
C       DURING THE 5-12 HOUR PERIOD.
C   -!- SIMILARITY: Contains 1 fork-head domain.
C   ---------------------------------------------------------------------
C   This SWISS-PROT entry is copyright. It is produced through a collaboration
C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
C   the European Bioinformatics Institute. There are no restrictions on its
C   use by non-profit institutions as long as its content is in no way
C   modified and this statement is not removed. Usage by and for commercial
C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
C   or send an email to license@isb-sib.ch).
C   ---------------------------------------------------------------------
R   EMBL; S80254; AAB35643.1; -.
R   EMBL; AE003594; AAF51727.1; -.
R   EMBL; M96440; AAF02177.1; -.
R   PIR; S59870; S59870.
R   HSSP; Q63245; 2HFH.
R   TRANSFAC; T02291; -.
R   FlyBase; FBgn0014143; croc.
R   GO; GO:0005634; C:nucleus; IDA.
R   InterPro; IPR001766; TF_Fork_head.
R   Pfam; PF00250; Fork_head; 1.
R   PRINTS; PR00053; FORKHEAD.
R   ProDom; PD000425; TF_Fork_head; 1.
R   SMART; SM00339; FH; 1.
R   PROSITE; PS00657; FORK_HEAD_1; 1.
R   PROSITE; PS00658; FORK_HEAD_2; 1.
R   PROSITE; PS50039; FORK_HEAD_3; 1.
W   DNA-binding; Developmental protein; Nuclear protein;
W   Transcription regulation.
T   DOMAIN       34     40       POLY-ALA.
T   DNA_BIND     69     160      FORK-HEAD.
T   DOMAIN      161     165      POLY-ARG.
T   DOMAIN      301     304      POLY-ALA.
T   DOMAIN      377     380      POLY-ASN.
T   DOMAIN      389     403      POLY-GLY.
T   DOMAIN      452     461      POLY-ALA.
T   DOMAIN      466     473      POLY-HIS.
T   VARIANT     122     122      L -> F (IN ALLELE CROC-75-3).
T   VARIANT     453     453      A -> V (IN ALLELE CROC-75-3).
Q   SEQUENCE   508 AA;  54516 MW;  2EFED1D8F63016D6 CRC64;

  Query Match             79.3%;  Score 23;  DB 1;  Length 508;
  Best Local Similarity   37.5%;  Pred. No. 4.4e+02;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              | | | |
Db        207 HMAAHAAH 214


RESULT 2
YM43_YEAST
ID   YM43_YEAST     STANDARD;      PRT;   510 AA.
AC   Q03218;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Hypothetical 56.2 kDa protein in SIP18-SPT21 intergenic region.
GN   YMR177W OR YM8010.07.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   PubMed=9169872;
RA   Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA   Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA   Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA   Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   XIII.";
RL   Nature 387:90-93(1997).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: STRONG, TO YEAST YPL224C.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z49808; CAA89910.1; -.
DR   PIR; S55124; S55124.
DR   SGD; S0004789; MMT1.
DR   GO; GO:0005739; C:mitochondrion; IDA.
DR   GO; GO:0006879; P:iron ion homeostasis; IGI.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM    165    185       POTENTIAL.
FT   TRANSMEM    194    214       POTENTIAL.
FT   TRANSMEM    241    261       POTENTIAL.
FT   TRANSMEM    286    306       POTENTIAL.
FT   TRANSMEM    333    353       POTENTIAL.
FT   TRANSMEM    356    376       POTENTIAL.
SQ   SEQUENCE   510 AA;  56209 MW;  F3CC9A230FB5DB87 CRC64;

  Query Match             79.3%;  Score 23;  DB 1;  Length 510;
  Best Local Similarity   37.5%;  Pred. No. 4.4e+02;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              | | | |
Db        137 HTHSHAAH 144


RESULT 3
BRC4_DROME
ID   BRC4_DROME     STANDARD;      PRT;   880 AA.
AC   Q24206; O46064; Q9W575;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
```

```
E   Broad-complex core-protein isoform 6.
N   BR OR BR-C OR EG:17A9.1 OR EG:25D2.1 OR EG:123F11.1 OR
N   CG11491/CG11514.
S   Drosophila melanogaster (Fruit fly).
C   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
C   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
C   Ephydroidea; Drosophilidae; Drosophila.
X   NCBI_TaxID=7227;
N   [1]
P   SEQUENCE FROM N.A., DEVELOPMENTAL STAGE, AND CHARACTERIZATION OF
P   ISOFORMS.
C   TISSUE=Imaginal disks, and Larva;
X   MEDLINE=96299417; PubMed=8660872;
A   Bayer C.A., Holley B., Fristrom J.W.;
T   "A switch in broad-complex zinc-finger isoform expression is regulated
T   posttranscriptionally during the metamorphosis of Drosophila imaginal
T   discs.";
L   Dev. Biol. 177:1-14(1996).
N   [2]
P   SEQUENCE FROM N.A.
C   STRAIN=Oregon-R;
X   MEDLINE=20196011; PubMed=10731137;
A   Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,
A   Barrell B.G., Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E.,
A   Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D.,
A   Minana B., Kafatos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,
A   Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablos B.,
A   Modolell J., Peter A., Schoettler P., Werner M., Mourkioti F.,
A   Beinert N., Dowe G., Schaefer U., Jaeckle H., Bucheton A.,
A   Callister D.M., Campbell L.A., Darlamitsou A., Henderson N.S.,
A   McMillan P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,
A   Glover D.M.;
T   "From sequence to chromosome: the tip of the X chromosome of D.
T   melanogaster.";
L   Science 287:2220-2222(2000).
N   [3]
P   SEQUENCE FROM N.A.
C   STRAIN=Berkeley;
X   MEDLINE=20196006; PubMed=10731132;
A   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
A   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
A   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
A   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
A   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
A   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
A   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
A   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
A   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
A   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
A   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
A   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
A   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
A   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
A   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
A   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
A   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
A   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
A   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
A   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
A   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
A   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
A   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
A   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
A   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
A   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
A   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
A   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
A   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
A   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
A   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
A   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
A   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
A   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA  Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA  Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT  "The genome sequence of Drosophila melanogaster.";
RL  Science 287:2185-2195(2000).
RN  [4]
RP  CHARACTERIZATION OF ISOFORMS, AND MUTATIONAL ANALYSIS.
RX  MEDLINE=97384928; PubMed=9242423;
RA  Bayer C.A., von Kalm L., Fristrom J.W.;
RT  "Relationships between protein isoforms and genetic functions
RT  demonstrate functional redundancy at the Broad-Complex during
RT  Drosophila metamorphosis.";
RL  Dev. Biol. 187:267-282(1997).
CC  -!- FUNCTION: BROAD-COMPLEX PROTEINS ARE REQUIRED FOR PUFFING AND
CC      TRANSCRIPTION OF SALIVARY GLAND LATE GENES DURING METAMORPHOSIS.
CC  -!- SUBCELLULAR LOCATION: Nuclear.
CC  -!- ALTERNATIVE PRODUCTS:
CC      Event=Alternative splicing; Named isoforms=6;
CC      Name=6; Synonyms=BCORE-Z4;
CC        IsoId=Q24206-1; Sequence=Displayed;
CC      Name=1; Synonyms=BCORE-TNT1-Q1-Z1;
CC        IsoId=Q01295-1; Sequence=External;
CC      Name=2; Synonyms=BCORE-Q1-Z1;
CC        IsoId=Q01295-2; Sequence=External;
CC      Name=3; Synonyms=BCORE-Q2-Z1;
CC        IsoId=Q01295-3; Sequence=External;
CC      Name=4; Synonyms=BCORE-Z2;
CC        IsoId=Q01295-4; Sequence=External;
CC      Name=5; Synonyms=BCORE-NS-Z3;
CC        IsoId=Q01295-5; Sequence=External;
CC  -!- DEVELOPMENTAL STAGE: ACCUMULATES TO A HIGH LEVEL AT THE BEGINNING
CC      OF THE ECDYSONE RESPONSE, DURING THE METAMORPHOSIS OF IMAGINAL
CC      DISKS IN PUFF STAGE 1, AND ABRUPTLY DISAPPEARS AFTER SEVERAL
CC      HOURS.
CC  -!- INDUCTION: INDUCED AS A PRIMARY RESPONSE TO 20-HYDROXYECDYSONE IN
CC      THIRD INSTAR LARVAL IMAGINAL DISKS.
CC  -!- SIMILARITY: Contains 1 BTB/POZ domain.
CC  -!- SIMILARITY: Contains 2 C2H2-type zinc fingers.
CC  -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 534
CC      TO 619 AND 656 TO 694 DUE TO FRAMESHIFTS.
CC  ---------------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  ---------------------------------------------------------------------------
DR  EMBL; U51585; AAB09760.1; ALT_FRAME.
DR  EMBL; AL009146; CAA15627.1; -.
DR  EMBL; AE003421; AAF45647.1; -.
DR  TRANSFAC; T01480; -.
DR  FlyBase; FBgn0000210; br.
DR  InterPro; IPR000210; BTB_POZ.
DR  InterPro; IPR007087; Znf_C2H2.
DR  Pfam; PF00651; BTB; 1.
DR  Pfam; PF00096; zf-C2H2; 2.
DR  SMART; SM00225; BTB; 1.
DR  SMART; SM00355; ZnF_C2H2; 2.
DR  PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.
DR  PROSITE; PS50157; ZINC_FINGER_C2H2_2; 2.
DR  PROSITE; PS50097; BTB; 1.
KW  Nuclear protein; DNA-binding; Developmental protein;
KW  Zinc-finger; Metal-binding; Alternative splicing.
FT  DOMAIN       32     97       BTB.
FT  ZN_FING     710    733       C2H2-TYPE 1.
FT  ZN_FING     740    763       C2H2-TYPE 2.
FT  DOMAIN      203    207       POLY-ALA.
FT  DOMAIN      265    268       POLY-ASN.
FT  DOMAIN      458    466       POLY-ASN.
FT  DOMAIN      584    589       POLY-PRO.
FT  DOMAIN      618    621       POLY-ALA.
FT  DOMAIN      798    803       POLY-ALA.
```

```
T   DOMAIN      821    833       POLY-ALA.
T   DOMAIN      862    867       POLY-GLN.
T   CONFLICT    436    436       G -> D (IN REF. 1).
T   CONFLICT    621    621       MISSING (IN REF. 1).
T   CONFLICT    624    624       A -> R (IN REF. 1).
T   CONFLICT    661    662       AV -> L (IN REF. 1).
T   CONFLICT    678    678       MISSING (IN REF. 1).
T   CONFLICT    722    723       KL -> NV (IN REF. 1).
Q   SEQUENCE   880 AA;  92305 MW;  500C0A4A38663AAF CRC64;

 Query Match             79.3%;  Score 23;  DB 1;  Length 880;
 Best Local Similarity   37.5%;  Pred. No. 7e+02;
 Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

y          1 HXXXHXXH 8
             |   |  |
b        625 HAHAHAAH 632


ESULT 4
YA_STRCO
D   SYA_STRCO      STANDARD;      PRT;   890 AA.
C   Q9KXP9;
T   28-FEB-2003 (Rel. 41, Created)
T   28-FEB-2003 (Rel. 41, Last sequence update)
T   28-FEB-2003 (Rel. 41, Last annotation update)
E   Alanyl-tRNA synthetase (EC 6.1.1.7) (Alanine--tRNA ligase) (AlaRS).
N   ALAS OR SCO1501 OR SC9C5.25C.
S   Streptomyces coelicolor.
C   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
C   Streptomycineae; Streptomycetaceae; Streptomyces.
X   NCBI_TaxID=1902;
N   [1]
P   SEQUENCE FROM N.A.
C   STRAIN=A3(2) / M145;
X   MEDLINE=21996410; PubMed=12000953;
A   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
A   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
A   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
A   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
A   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
A   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
A   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
A   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
A   Hopwood D.A.;
T   "Complete genome sequence of the model actinomycete Streptomyces
T   coelicolor A3(2).";
L   Nature 417:141-147(2002).
C   -!- CATALYTIC ACTIVITY: ATP + L-alanine + tRNA(Ala) = AMP +
C       diphosphate + L-alanyl-tRNA(Ala).
C   -!- SUBCELLULAR LOCATION: Cytoplasmic.
C   -!- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
C   ------------------------------------------------------------------------
C   This SWISS-PROT entry is copyright. It is produced through a collaboration
C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
C   the European Bioinformatics Institute. There are no restrictions on its
C   use by non-profit institutions as long as its content is in no way
C   modified and this statement is not removed. Usage by and for commercial
C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
C   or send an email to license@isb-sib.ch).
C   ------------------------------------------------------------------------
R   EMBL; AL939109; CAB93381.1; -.
R   HAMAP; MF_00036; -; 1.
R   InterPro; IPR003156; DHHA1.
R   InterPro; IPR002318; tRNA-synt_2c.
R   InterPro; IPR006193; tRNA_synt_Ala.
R   Pfam; PF02272; DHHA1; 1.
R   Pfam; PF01411; tRNA-synt_2c; 1.
R   PRINTS; PR00980; TRNASYNTHALA.
R   TIGRFAMs; TIGR00344; alaS; 1.
R   PROSITE; PS50860; AA_TRNA_LIGASE_II_ALA; 1.
W   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
W   Complete proteome.

SQ   SEQUENCE   890 AA;  95786 MW;  05B2FD563D35F4DF CRC64;

  Query Match             79.3%;  Score 23;  DB 1;  Length 890;
  Best Local Similarity   37.5%;  Pred. No. 7.1e+02;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db        573 HSATHLTH 580


RESULT 5
GSR1_HUMAN
ID   GSR1_HUMAN      STANDARD;      PRT;  1509 AA.
AC   Q9N2M4;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Glioma tumor suppressor candidate region gene 1 protein.
GN   GLTSCR1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX   MEDLINE=20175430; PubMed=10708517;
RA   Smith J.S., Tachibana I., Pohl U., Lee H.K., Thanarajasingam U.,
RA   Portier B.P., Ueki K., Billings S., Ramaswamy S., Mohrenweiser H.W.,
RA   Scheithauer B.W., Louis D.N., Jenkins R.B.;
RT   "A transcript map of the chromosome 19q-Arm glioma tumor suppressor
RT   region.";
RL   Genomics 64:44-50(2000).
CC   -!- TISSUE SPECIFICITY: Expressed at moderate levels in heart, brain,
CC       placenta, skeletal muscle, and pancreas, and at lower levels in
CC       lung, liver, and kidney.
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; AF182077; AAF62874.1; -.
DR   Genew; HGNC:4332; GLTSCR1.
DR   MIM; 605690; -.
FT   DOMAIN       37     45       POLY-GLY.
FT   DOMAIN      884    889       POLY-PRO.
FT   DOMAIN     1214   1225       POLY-SER.
FT   DOMAIN     1282   1286       POLY-PRO.
FT   DOMAIN     1294   1304       POLY-PRO.
SQ   SEQUENCE   1509 AA;  152991 MW;  7C5144F443CE6821 CRC64;

  Query Match             79.3%;  Score 23;  DB 1;  Length 1509;
  Best Local Similarity   37.5%;  Pred. No. 1.1e+03;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db        482 HSGAHSAH 489


RESULT 6
YN78_YEAST
ID   YN78_YEAST      STANDARD;      PRT;    52 AA.
AC   P53820;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical 6.0 kDa protein in COS1 5'region.
GN   YNL338W OR N0170.
```

```
S   Saccharomyces cerevisiae (Baker's yeast).
C   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
C   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
X   NCBI_TaxID=4932;
N   [1]
P   SEQUENCE FROM N.A.
A   Obermaier B., Piravandi E., Rinke M.;
L   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
C   -!- SIMILARITY: TO YEAST YHR217C.
C   ---------------------------------------------------------------------
C   This SWISS-PROT entry is copyright. It is produced through a collaboration
C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
C   the European Bioinformatics Institute. There are no restrictions on its
C   use by non-profit institutions as long as its content is in no way
C   modified and this statement is not removed. Usage by and for commercial
C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
C   or send an email to license@isb-sib.ch).
C   ---------------------------------------------------------------------
R   EMBL; Z71614; CAA96274.1; -.
R   EMBL; Z71613; CAA96273.1; -.
R   PIR; S63324; S63324.
R   SGD; S0005282; YNL338W.
W   Hypothetical protein.
Q   SEQUENCE   52 AA;  5951 MW;  C1E4066D43E057A1 CRC64;

 Query Match             75.9%;  Score 22;  DB 1;  Length 52;
 Best Local Similarity   37.5%;  Pred. No. 90;
 Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

y          1 HXXXHXXH 8
               |   |  |
b         38 HTHTHTHH 45


ESULT 7
034_TREPA
D   Y034_TREPA     STANDARD;      PRT;   316 AA.
C   O83077;
T   16-OCT-2001 (Rel. 40, Created)
T   16-OCT-2001 (Rel. 40, Last sequence update)
T   15-SEP-2003 (Rel. 42, Last annotation update)
E   Putative periplasmic metal-binding protein TP0034 precursor.
N   TP0034.
S   Treponema pallidum.
C   Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
X   NCBI_TaxID=160;
N   [1]
P   SEQUENCE FROM N.A.
C   STRAIN=Nichols;
X   MEDLINE=98332770; PubMed=9665876;
A   Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
A   Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
A   Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
A   Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
A   McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
A   Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
A   Venter J.C.;
T   "Complete genome sequence of Treponema pallidum, the syphilis
T   spirochete.";
L   Science 281:375-388(1998).
C   -!- FUNCTION: PART OF AN ATP-DRIVEN TRANSPORT SYSTEM
C       TP0034/TP0035/TP0036 FOR A METAL. METAL-BINDING COMPONENT.
C   -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
C   -!- SIMILARITY: BELONGS TO THE BACTERIAL SOLUTE-BINDING PROTEIN FAMILY
C       9.
C   ---------------------------------------------------------------------
C   This SWISS-PROT entry is copyright. It is produced through a collaboration
C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
C   the European Bioinformatics Institute. There are no restrictions on its
C   use by non-profit institutions as long as its content is in no way
C   modified and this statement is not removed. Usage by and for commercial
C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
C   or send an email to license@isb-sib.ch).
CC  ---------------------------------------------------------------------
DR  EMBL; AE001188; AAC65029.1; -.
DR  PIR; D71375; D71375.
DR  TIGR; TP0034; -.
DR  InterPro; IPR006128; Lipoprotein_4.
DR  InterPro; IPR006127; SBP_bac_9.
DR  Pfam; PF01297; SBP_bac_9; 1.
DR  PRINTS; PR00690; ADHESNFAMILY.
DR  PROSITE; PS00013; PROKAR_LIPOPROTEIN; UNKNOWN_1.
KW  Hypothetical protein; Transport; Periplasmic; Metal-binding; Signal;
KW  Complete proteome.
FT  SIGNAL        1     19       POTENTIAL.
FT  CHAIN        20    316       PUTATIVE PERIPLASMIC METAL-BINDING
FT                               PROTEIN TP0034.
SQ  SEQUENCE   316 AA;  35433 MW;  16051C2199BC81AB CRC64;

 Query Match             75.9%;  Score 22;  DB 1;  Length 316;
 Best Local Similarity   37.5%;  Pred. No. 4.4e+02;
 Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
                |   |  |
Db        124 HTRGHTAH 131


RESULT 8
ADH1_BACST
ID  ADH1_BACST     STANDARD;      PRT;   337 AA.
AC  P12311;
DT  01-OCT-1989 (Rel. 12, Created)
DT  01-NOV-1995 (Rel. 32, Last sequence update)
DT  15-SEP-2003 (Rel. 42, Last annotation update)
DE  Alcohol dehydrogenase (EC 1.1.1.1) (ADH-T).
GN  ADHT.
OS  Bacillus stearothermophilus.
OC  Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.
OX  NCBI_TaxID=1422;
RN  [1]
RP  SEQUENCE FROM N.A., AND MUTAGENESIS.
RC  STRAIN=NCA 1503;
RX  MEDLINE=92138636; PubMed=1735726;
RA  Sakoda H., Imanaka T.;
RT  "Cloning and sequencing of the gene coding for alcohol dehydrogenase
RT  of Bacillus stearothermophilus and rational shift of the optimum
RT  pH.";
RL  J. Bacteriol. 174:1397-1402(1992).
RN  [2]
RP  SEQUENCE OF 1-45.
RX  MEDLINE=73229257; PubMed=4578954;
RA  Bridgen J., Kolb E., Harris J.I.;
RT  "Amino acid sequence homology in alcohol dehydrogenase.";
RL  FEBS Lett. 33:1-3(1973).
RN  [3]
RP  SEQUENCE OF 34-54.
RX  MEDLINE=79169263; PubMed=436831;
RA  Jeck R., Woenckhaus C., Harris J.I., Runswick M.J.;
RT  "Identification of the amino acid residue modified in Bacillus
RT  stearothermophilus alcohol dehydrogenase by the NAD+ analogue 4-(3-
RT  bromoacetylpyridinio)butyldiphosphoadenosine.";
RL  Eur. J. Biochem. 93:57-64(1979).
RN  [4]
RP  SEQUENCE OF 1-37; 188-197; 247-263 AND 324-336.
RC  STRAIN=NCA 1503;
RX  MEDLINE=94325354; PubMed=8049268;
RA  Robinson G.A., Bailey C.J., Dowds B.C.A.;
RT  "Gene structure and amino acid sequences of alcohol dehydrogenases of
RT  Bacillus stearothermophilus.";
RL  Biochim. Biophys. Acta 1218:432-434(1994).
CC  -!- FUNCTION: THERMOSTABLE NAD(+)-DEPENDENT ALCOHOL DEHYDROGENASE.
CC  -!- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +
CC      NADH.
CC  -!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
CC  -!- ENZYME REGULATION: SUBSTRATE INHIBITION IS NOT OBSERVED WITH ANY
```

```
C       ALCOHOLS, AND THE ENZYME-NADH DISSOCIATION IS NOT CONSIDERED TO BE
C       A RATE-LIMITING STEP.
C    -!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
C       family.
C    ---------------------------------------------------------------------
C    This SWISS-PROT entry is copyright. It is produced through a collaboration
C    between the Swiss Institute of Bioinformatics and the EMBL outstation -
C    the European Bioinformatics Institute. There are no restrictions on its
C    use by non-profit institutions as long as its content is in no way
C    modified and this statement is not removed. Usage by and for commercial
C    entities requires a license agreement (See http://www.isb-sib.ch/announce/
C    or send an email to license@isb-sib.ch).
C    ---------------------------------------------------------------------
R    EMBL; D90421; BAA14411.1; -.
R    PIR; A42654; A42654.
R    InterPro; IPR002328; ADH_zinc.
R    InterPro; IPR002085; Adh_zn_family.
R    Pfam; PF00107; ADH_zinc_N; 1.
R    PROSITE; PS00059; ADH_ZINC; 1.
W    Oxidoreductase; Zinc; Metal-binding; NAD.
T    METAL        38     38       ZINC 1 (CATALYTIC) (BY SIMILARITY).
T    METAL        61     61       ZINC 1 (CATALYTIC) (BY SIMILARITY).
T    METAL        92     92       ZINC 2 (BY SIMILARITY).
T    METAL        95     95       ZINC 2 (BY SIMILARITY).
T    METAL        98     98       ZINC 2 (BY SIMILARITY).
T    METAL       106    106       ZINC 2 (BY SIMILARITY).
T    METAL       148    148       ZINC 1 (CATALYTIC) (BY SIMILARITY).
T    MUTAGEN      40     40       T->S: LITTLE DECREASE IN ACTIVITY.
T    MUTAGEN      43     43       H->R: HIGHER LEVEL OF ACTIVITY AT PH 9.
T    MUTAGEN      38     38       C->S: NO ACTIVITY.
T    MUTAGEN      40     40       T->A: NO ACTIVITY.
T    MUTAGEN      43     43       H->A: NO ACTIVITY.
T    CONFLICT     22     22       MISSING (IN REF. 2).
T    CONFLICT     33     33       MISSING (IN REF. 2).
T    CONFLICT     52     53       KP -> PK (IN REF. 3).
Q    SEQUENCE   337 AA;  36100 MW;  B9B35A80EE9B7A86 CRC64;

 Query Match             75.9%;  Score 22;  DB 1;  Length 337;
 Best Local Similarity   37.5%;  Pred. No. 4.6e+02;
 Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db         39 HTDLHAAH 46


RESULT 9
ADH2_BACST
ID   ADH2_BACST     STANDARD;      PRT;   339 AA.
AC   P42327;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Alcohol dehydrogenase (EC 1.1.1.1) (ADH).
GN   ADH.
OS   Bacillus stearothermophilus.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.
OX   NCBI_TaxID=1422;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-40.
RC   STRAIN=DSM 2334;
RX   MEDLINE=94325354; PubMed=8049268;
RA   Robinson G.A., Bailey C.J., Dowds B.C.A.;
RT   "Gene structure and amino acid sequences of alcohol dehydrogenases of
RT   Bacillus stearothermophilus.";
RL   Biochim. Biophys. Acta 1218:432-434(1994).
CC   -!- FUNCTION: ACTIVE WITH PRIMARY ALCOHOLS, INCLUDING METHANOL.
CC   -!- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +
CC       NADH.
CC   -!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
CC   -!- ENZYME REGULATION: THE RATE-LIMITING STEP IS NADH RELEASE.
CC       CATABOLITE REPRESSION.
CC   -!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
CC       family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z25544; CAA80989.1; -.
DR   PIR; S47643; S47643.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   Pfam; PF00107; ADH_zinc_N; 1.
DR   PROSITE; PS00059; ADH_ZINC; 1.
KW   Oxidoreductase; Zinc; Metal-binding; NAD.
FT   METAL        38     38       ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT   METAL        61     61       ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT   METAL        92     92       ZINC 2 (BY SIMILARITY).
FT   METAL        95     95       ZINC 2 (BY SIMILARITY).
FT   METAL        98     98       ZINC 2 (BY SIMILARITY).
FT   METAL       106    106       ZINC 2 (BY SIMILARITY).
FT   METAL       148    148       ZINC 1 (CATALYTIC) (BY SIMILARITY).
SQ   SEQUENCE   339 AA;  36205 MW;  0EC33CE7287D7476 CRC64;

  Query Match             75.9%;  Score 22;  DB 1;  Length 339;
  Best Local Similarity   37.5%;  Pred. No. 4.6e+02;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db         39 HTDLHAAH 46


RESULT 10
ADH3_BACST
ID   ADH3_BACST     STANDARD;      PRT;   339 AA.
AC   P42328;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Alcohol dehydrogenase (EC 1.1.1.1) (ADH-HT).
OS   Bacillus stearothermophilus.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.
OX   NCBI_TaxID=1422;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCIMB 12403 / LLD-R;
RX   MEDLINE=94291628; PubMed=8020473;
RA   Cannio R., Rossi M., Bartolucci S.;
RT   "A few amino acid substitutions are responsible for the higher
RT   thermostability of a novel NAD(+)-dependent bacillar alcohol
RT   dehydrogenase.";
RL   Eur. J. Biochem. 222:345-352(1994).
CC   -!- FUNCTION: THERMOSTABLE AND THERMOPHILIC NAD(+)-DEPENDENT
CC       ALCOHOL DEHYDROGENASE. BEARS MAINLY AN ETHANOL-DEHYDROGENASE
CC       ACTIVITY.
CC   -!- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +
CC       NADH.
CC   -!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
CC   -!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
CC       family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z27089; CAA81612.1; -.
```

```
R   PIR; S45605; S45605.
R   HSSP; P28304; 1QOR.
R   InterPro; IPR002328; ADH_zinc.
R   InterPro; IPR002085; Adh_zn_family.
R   Pfam; PF00107; ADH_zinc_N; 1.
R   PROSITE; PS00059; ADH_ZINC; 1.
W   Oxidoreductase; Zinc; Metal-binding; NAD.
T   METAL        38     38       ZINC 1 (CATALYTIC) (BY SIMILARITY).
T   METAL        61     61       ZINC 1 (CATALYTIC) (BY SIMILARITY).
T   METAL        92     92       ZINC 2 (BY SIMILARITY).
T   METAL        95     95       ZINC 2 (BY SIMILARITY).
T   METAL        98     98       ZINC 2 (BY SIMILARITY).
T   METAL       106    106       ZINC 2 (BY SIMILARITY).
T   METAL       148    148       ZINC 1 (CATALYTIC) (BY SIMILARITY).
Q   SEQUENCE   339 AA;  36338 MW;  AED17E4A34163430 CRC64;

 Query Match             75.9%;  Score 22;  DB 1;  Length 339;
 Best Local Similarity   37.5%;  Pred. No. 4.6e+02;
 Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

y          1 HXXXHXXH 8
               |   |  |
b         39 HTDLHAAH 46


ESULT 11
ILA_HUMAN
D   FILA_HUMAN      STANDARD;      PRT;   416 AA.
C   P20930;
T   01-FEB-1991 (Rel. 17, Created)
T   01-FEB-1996 (Rel. 33, Last sequence update)
T   28-FEB-2003 (Rel. 41, Last annotation update)
E   Filaggrin precursor (Fragment).
N   FLG.
S   Homo sapiens (Human).
C   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
X   NCBI_TaxID=9606;
N   [1]
P   SEQUENCE FROM N.A.
X   MEDLINE=89296901; PubMed=2740331;
A   McKinley-Grant L.J., Idler W.W., Bernstein I.A., Parry D.A.D.,
A   Cannizzaro L., Croce C.M., Huebner K., Lessin S.R., Steinert P.M.;
T   "Characterization of a cDNA clone encoding human filaggrin and
T   localization of the gene to chromosome region 1q21.";
L   Proc. Natl. Acad. Sci. U.S.A. 86:4848-4852(1989).
N   [2]
P   CITRULLINATION.
X   MEDLINE=96374388; PubMed=8780679;
A   Senshu T., Kan S., Ogawa H., Manabe M., Asaga H.;
T   "Preferential deimination of keratin K1 and filaggrin during the
T   terminal differentiation of human epidermis.";
L   Biochem. Biophys. Res. Commun. 225:712-719(1996).
C   -!- FUNCTION: AGGREGATES KERATIN INTERMEDIATE FILAMENTS AND PROMOTES
C       DISULFIDE-BOND FORMATION AMONG THE INTERMEDIATE FILAMENTS DURING
C       TERMINAL DIFFERENTIATION OF MAMMALIAN EPIDERMIS.
C   -!- PTM: FILAGGRIN IS INITIALLY SYNTHESIZED AS A LARGE, INSOLUBLE,
C       HIGHLY PHOSPHORYLATED PRECURSOR CONTAINING MANY TANDEM COPIES
C       OF 324 AA, WHICH ARE NOT SEPARATED BY "LARGE LINKER". THE
C       PRECURSOR IS DEPOSITED AS KERATOHYALIN GRANULES. DURING TERMINAL
C       DIFFERENTIATION IT IS DEPHOSPHORYLATED & PROTEOLYTICALLY CLEAVED.
C   -!- PTM: Undergoes deimination of some arginine residues
C       (citrullination).
C   ---------------------------------------------------------------------
C   This SWISS-PROT entry is copyright. It is produced through a collaboration
C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
C   the European Bioinformatics Institute. There are no restrictions on its
C   use by non-profit institutions as long as its content is in no way
C   modified and this statement is not removed. Usage by and for commercial
C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
C   or send an email to license@isb-sib.ch).
C   ---------------------------------------------------------------------
R   EMBL; M24355; AAA52454.1; -.
DR  PIR; A32947; A32947.
DR  Genew; HGNC:3748; FLG.
DR  MIM; 135940; -.
DR  GO; GO:0005882; C:intermediate filament; NAS.
DR  GO; GO:0005198; F:structural molecule activity; NAS.
DR  GO; GO:0007275; P:development; NAS.
DR  InterPro; IPR003303; Filaggrin.
DR  Pfam; PF03516; Filaggrin; 2.
DR  PRINTS; PR00487; FILAGGRIN.
KW  Phosphorylation; Citrullination; Developmental protein.
FT  NON_TER      1      1
SQ  SEQUENCE   416 AA;  44105 MW;  DEEA3218BA043F32 CRC64;

  Query Match             75.9%;  Score 22;  DB 1;  Length 416;
  Best Local Similarity   37.5%;  Pred. No. 5.6e+02;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
                |   |  |
Db         99 HSGSHHSH 106


RESULT 12
PFTB_PEA
ID   PFTB_PEA       STANDARD;      PRT;   419 AA.
AC   Q04903;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Protein farnesyltransferase beta subunit (EC 2.5.1.-) (CAAX
DE   farnesyltransferase beta subunit) (RAS proteins prenyltransferase
DE   beta) (PTase-beta).
GN   FTB.
OS   Pisum sativum (Garden pea).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Vicieae; Pisum.
OX   NCBI_TaxID=3888;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Alaska; TISSUE=Seedling;
RX   MEDLINE=94105305; PubMed=8278509;
RA   Yang Z., Cramer C.L., Watson J.C.;
RT   "Protein farnesyltransferase in plants. Molecular cloning and
RT   expression of a homolog of the beta subunit from the garden pea.";
RL   Plant Physiol. 101:667-674(1993).
CC   -!- FUNCTION: CATALYZES THE TRANSFER OF A FARNESYL MOIETY FROM
CC       FARNESYL PYROPHOSPHATE TO A CYSTEINE AT THE FOURTH POSITION FROM
CC       THE C-TERMINUS OF SEVERAL PROTEINS. THE BETA SUBUNIT IS
CC       RESPONSIBLE FOR PEPTIDE-BINDING (BY SIMILARITY).
CC   -!- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC   -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE PROTEIN PRENYLTRANSFERASE BETA SUBUNIT
CC       FAMILY.
CC   -!- SIMILARITY: Contains 5 PFTB repeats.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L08664; AAA33649.1; -.
DR   PIR; JQ2254; JQ2254.
DR   HSSP; Q02293; 1FT1.
DR   InterPro; IPR001330; Prenyltrans.
DR   Pfam; PF00432; prenyltrans; 5.
KW   Transferase; Prenyltransferase; Repeat; Zinc.
FT   REPEAT       68    109       PFTB 1.
FT   REPEAT      119    160       PFTB 2.
```

```
T   REPEAT      167    208       PFTB 3.
T   REPEAT      215    256       PFTB 4.
T   REPEAT      329    371       PFTB 5.
T   METAL       241    241       ZINC (BY SIMILARITY).
T   METAL       243    243       ZINC (BY SIMILARITY).
T   METAL       359    359       ZINC (BY SIMILARITY).
Q   SEQUENCE   419 AA;  46793 MW;  4F040E0094277D7C CRC64;

 Query Match             75.9%;  Score 22;  DB 1;  Length 419;
 Best Local Similarity   37.5%;  Pred. No. 5.6e+02;
 Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

y          1 HXXXHXXH 8
             |   |  |
b        291 HATSHIRH 298


ESULT 13
ROA_NEIMB
D   PROA_NEIMB     STANDARD;      PRT;   420 AA.
C   Q9JZG3;
T   16-OCT-2001 (Rel. 40, Created)
T   16-OCT-2001 (Rel. 40, Last sequence update)
T   28-FEB-2003 (Rel. 41, Last annotation update)
E   Gamma-glutamyl phosphate reductase (GPR) (EC 1.2.1.41) (Glutamate-5-
E   semialdehyde dehydrogenase) (Glutamyl-gamma-semialdehyde
E   dehydrogenase) (GSA dehydrogenase).
N   PROA OR NMB1068.
S   Neisseria meningitidis (serogroup B).
C   Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
C   Neisseriaceae; Neisseria.
X   NCBI_TaxID=491;
N   [1]
P   SEQUENCE FROM N.A.
C   STRAIN=MC58 / Serogroup B;
X   MEDLINE=20175755; PubMed=10710307;
A   Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
A   Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
A   Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
A   Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
A   Mason T., Ciecko A., Parksey D.S., Blair E., Cittone H., Clark E.B.,
A   Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
A   Gill J., Scarlato V., Masignani V., Pizza M., Grandi G., Sun L.,
A   Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
T   "Complete genome sequence of Neisseria meningitidis serogroup B strain
T   MC58.";
L   Science 287:1809-1815(2000).
C   -!- FUNCTION: CATALYZES THE NADPH DEPENDENT REDUCTION OF L-GAMMA-
C       GLUTAMYL 5-PHOSPHATE INTO L-GLUTAMATE 5-SEMIALDEHYDE AND
C       PHOSPHATE. THE PRODUCT SPONTANEOUSLY UNDERGOES CYCLIZATION TO FORM
C       1-PYRROLINE-5-CARBOXYLATE.
C   -!- CATALYTIC ACTIVITY: L-glutamate 5-semialdehyde + phosphate +
C       NADP(+) = L-gamma-glutamyl 5-phosphate + NADPH.
C   -!- PATHWAY: Proline biosynthesis; second step.
C   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
C   -!- SIMILARITY: BELONGS TO THE GAMMA-GLUTAMYL PHOSPHATE REDUCTASE
C       FAMILY.
C   ---------------------------------------------------------------------
C   This SWISS-PROT entry is copyright. It is produced through a collaboration
C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
C   the European Bioinformatics Institute. There are no restrictions on its
C   use by non-profit institutions as long as its content is in no way
C   modified and this statement is not removed. Usage by and for commercial
C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
C   or send an email to license@isb-sib.ch).
C   ---------------------------------------------------------------------
R   EMBL; AE002457; AAF62324.1; -.
R   TIGR; NMB1068; -.
R   HAMAP; MF_00412; -; 1.
R   InterPro; IPR002086; Aldehyde_dehydr.
R   InterPro; IPR000965; Gglut_pp_reduct.
R   Pfam; PF00171; aldedh; 1.
R   TIGRFAMs; TIGR00407; proA; 1.
R   PROSITE; PS01223; PROA; 1.
W   Oxidoreductase; Proline biosynthesis; NADP; Complete proteome.
Q   SEQUENCE   420 AA;  45256 MW;  009996E9CF6B1118 CRC64;

  Query Match             75.9%;  Score 22;  DB 1;  Length 420;
  Best Local Similarity   37.5%;  Pred. No. 5.6e+02;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy         1 HXXXHXXH 8
             |   |  |
Db       335 HIETHSTH 342


RESULT 14
PROA_STRPN
ID   PROA_STRPN     STANDARD;      PRT;   420 AA.
AC   Q97R94;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Gamma-glutamyl phosphate reductase (GPR) (EC 1.2.1.41) (Glutamate-5-
DE   semialdehyde dehydrogenase) (Glutamyl-gamma-semialdehyde
DE   dehydrogenase) (GSA dehydrogenase).
GN   PROA OR SP0932.
OS   Streptococcus pneumoniae.
OC   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1313;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC BAA-334 / TIGR4;
RX   MEDLINE=21357209; PubMed=11463916;
RA   Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA   Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA   Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA   Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA   Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA   McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA   Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA   Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT   "Complete genome sequence of a virulent isolate of Streptococcus
RT   pneumoniae.";
RL   Science 293:498-506(2001).
CC   -!- FUNCTION: Catalyzes the NADPH dependent reduction of L-gamma-
CC       glutamyl 5-phosphate into L-glutamate 5-semialdehyde and
CC       phosphate. The product spontaneously undergoes cyclization to form
CC       1-pyrroline-5-carboxylate.
CC   -!- CATALYTIC ACTIVITY: L-glutamate 5-semialdehyde + phosphate +
CC       NADP(+) = L-gamma-glutamyl 5-phosphate + NADPH.
CC   -!- PATHWAY: Proline biosynthesis; second step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE GAMMA-GLUTAMYL PHOSPHATE REDUCTASE
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE007398; AAK75056.1; -.
DR   PIR; G95107; G95107.
DR   TIGR; SP0932; -.
DR   HAMAP; MF_00412; -; 1.
DR   InterPro; IPR002086; Aldehyde_dehydr.
DR   InterPro; IPR000965; Gglut_pp_reduct.
DR   Pfam; PF00171; aldedh; 1.
DR   TIGRFAMs; TIGR00407; proA; 1.
DR   PROSITE; PS01223; PROA; 1.
KW   Oxidoreductase; Proline biosynthesis; NADP; Complete proteome.
SQ   SEQUENCE   420 AA;  45241 MW;  A5D96CEDE50E87A2 CRC64;
```

```
 Query Match             75.9%;  Score 22;  DB 1;  Length 420;
 Best Local Similarity   37.5%;  Pred. No. 5.6e+02;
 Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

)y          1 HXXXHXXH 8
              |   |  |
)b        335 HIESHSTH 342


'ESULT 15
'KH2_XENLA
 D   FKH2_XENLA     STANDARD;      PRT;   427 AA.
 C   P32315;
 T   01-OCT-1993 (Rel. 27, Created)
 T   01-OCT-1993 (Rel. 27, Last sequence update)
 T   01-NOV-1995 (Rel. 32, Last annotation update)
 E   XFKH2 protein.
 N   XFKH2.
 S   Xenopus laevis (African clawed frog).
 C   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 C   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
 C   Xenopodinae; Xenopus.
 X   NCBI_TaxID=8355;
 N   [1]
 P   SEQUENCE FROM N.A.
 X   MEDLINE=94074768; PubMed=8253274;
 A   Bolce M.E., Hemmati-Brivanlou A., Harland R.M.;
 T   "XFKH2, a Xenopus HNF-3 alpha homologue, exhibits both
 T   activin-inducible and autonomous phases of expression in early
 T   embryos.";
 L   Dev. Biol. 160:413-423(1993).
 C   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 C   -!- TISSUE SPECIFICITY: PRESENT IN THE VEGETAL POLE AND MARGINAL ZONE
 C       BUT NOT THE ANIMAL POLE OF GASTRULAE AND IN EQUAL LEVELS IN THE
 C       DORSAL AND VENTRAL HALVES OF BOTH GASTRULAE AND NEURULAE.
 C   -!- SIMILARITY: Contains 1 fork-head domain.
 C   ---------------------------------------------------------------------
 C   This SWISS-PROT entry is copyright. It is produced through a collaboration
 C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
 C   the European Bioinformatics Institute. There are no restrictions on its
 C   use by non-profit institutions as long as its content is in no way
 C   modified and this statement is not removed. Usage by and for commercial
 C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
 C   or send an email to license@isb-sib.ch).
 C   ---------------------------------------------------------------------
 R   EMBL; M93658; AAA17050.1; -.
 R   PIR; I51580; I51580.
 R   HSSP; Q63245; 2HFH.
 R   InterPro; IPR001766; TF_Fork_head.
 R   Pfam; PF00250; Fork_head; 1.
 R   PRINTS; PR00053; FORKHEAD.
 R   ProDom; PD000425; TF_Fork_head; 1.
 R   SMART; SM00339; FH; 1.
 R   PROSITE; PS00657; FORK_HEAD_1; 1.
 R   PROSITE; PS00658; FORK_HEAD_2; 1.
 R   PROSITE; PS50039; FORK_HEAD_3; 1.
 W   DNA-binding; Nuclear protein.
 T   DNA_BIND    156    247       FORK-HEAD.
 Q   SEQUENCE   427 AA;  46572 MW;  2D29A42AF960730C CRC64;

  Query Match             75.9%;  Score 22;  DB 1;  Length 427;
  Best Local Similarity   37.5%;  Pred. No. 5.7e+02;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

)y          1 HXXXHXXH 8
              |   |  |
)b        333 HSLAHETH 340


!earch completed: November 21, 2003, 15:48:33
!ob time : 10 secs
```

GenCore version 5.1.6


M protein - protein search, using sw model

un on: November 21, 2003, 15:43:40 ; Search time 28 Seconds
(without alignments)
73.729 Million cell updates/sec

itle: US-10-064-903-1
erfect score: 29
equence: 1 HXXXHXXH 8

coring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

earched: 830525 seqs, 258052604 residues

otal number of hits satisfying chosen parameters: 830525

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| esult No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 23 | 79.3 | 119 | 16 | Q92K02 | Q92k02 rhizobium m |
| 2 | 23 | 79.3 | 152 | 17 | Q9YE47 | Q9ye47 aeropyrum p |
| 3 | 23 | 79.3 | 177 | 5 | Q9XWQ9 | Q9xwq9 caenorhabdi |
| 4 | 23 | 79.3 | 204 | 2 | Q9WX65 | Q9wx65 acetobacter |
| 5 | 23 | 79.3 | 240 | 16 | Q9PFU7 | Q9pfu7 xylella fas |
| 6 | 23 | 79.3 | 324 | 2 | Q9F502 | Q9f502 anabaena va |
| 7 | 23 | 79.3 | 327 | 2 | Q9FD42 | Q9fd42 anabaena sp |
| 8 | 23 | 79.3 | 327 | 16 | Q8YU45 | Q8yu45 anabaena sp |
| 9 | 23 | 79.3 | 341 | 5 | Q8T468 | Q8t468 drosophila |
| 10 | 23 | 79.3 | 341 | 5 | Q8MY62 | Q8my62 ciona savig |
| 11 | 23 | 79.3 | 342 | 5 | Q18888 | Q18888 caenorhabdi |
| 12 | 23 | 79.3 | 382 | 16 | O54173 | O54173 streptomyce |
| 13 | 23 | 79.3 | 402 | 2 | Q9KX27 | Q9kx27 oligotropha |
| 14 | 23 | 79.3 | 412 | 13 | Q98UI5 | Q98ui5 lampetra ja |
| 15 | 23 | 79.3 | 466 | 10 | Q947K4 | Q947k4 brassica na |
| 16 | 23 | 79.3 | 472 | 5 | Q09657 | Q09657 caenorhabdi |
| 17 | 23 | 79.3 | 534 | 5 | Q8MTV4 | Q8mtv4 plasmodium |
| 18 | 23 | 79.3 | 536 | 5 | Q8I3N3 | Q8i3n3 plasmodium |
| 19 | 23 | 79.3 | 596 | 16 | Q9L1Q1 | Q9l1q1 streptomyce |
| 20 | 23 | 79.3 | 606 | 16 | O31566 | O31566 bacillus su |
| 21 | 23 | 79.3 | 610 | 2 | O52961 | O52961 bacillus su |
| 22 | 23 | 79.3 | 635 | 4 | Q13476 | Q13476 homo sapien |
| 23 | 23 | 79.3 | 676 | 5 | Q95WV0 | Q95wv0 drosophila |
| 24 | 23 | 79.3 | 677 | 6 | Q28256 | Q28256 canis famil |
| 25 | 23 | 79.3 | 690 | 5 | Q8IM21 | Q8im21 plasmodium |
| 26 | 23 | 79.3 | 696 | 5 | Q9VDP3 | Q9vdp3 drosophila |
| 27 | 23 | 79.3 | 713 | 16 | Q8XR50 | Q8xr50 ralstonia s |
| 28 | 23 | 79.3 | 826 | 10 | Q9SCU4 | Q9scu4 arabidopsis |
| 29 | 23 | 79.3 | 826 | 10 | Q9SCU3 | Q9scu3 arabidopsis |
| 30 | 23 | 79.3 | 899 | 3 | Q8NIZ0 | Q8niz0 neurospora |
| 31 | 23 | 79.3 | 1031 | 5 | Q8MT64 | Q8mt64 drosophila |
| 32 | 23 | 79.3 | 1059 | 5 | Q9VZ52 | Q9vz52 drosophila |
| 33 | 23 | 79.3 | 1182 | 5 | Q9VXL1 | Q9vxl1 drosophila |
| 34 | 23 | 79.3 | 1226 | 5 | Q9V4U3 | Q9v4u3 drosophila |
| 35 | 23 | 79.3 | 1359 | 5 | Q9VX26 | Q9vx26 drosophila |
| 36 | 23 | 79.3 | 3036 | 4 | Q8TDJ6 | Q8tdj6 homo sapien |
| 37 | 23 | 79.3 | 3469 | 5 | Q9U4I2 | Q9u4i2 drosophila |
| 38 | 23 | 79.3 | 3604 | 5 | Q9VYK0 | Q9vyk0 drosophila |
| 39 | 23 | 79.3 | 4360 | 3 | Q9UVN5 | Q9uvn5 alternaria |
| 40 | 22 | 75.9 | 56 | 3 | Q8TGJ7 | Q8tgj7 saccharomyc |
| 41 | 22 | 75.9 | 57 | 16 | Q8PE57 | Q8pe57 xanthomonas |
| 42 | 22 | 75.9 | 61 | 16 | Q8ZE32 | Q8ze32 yersinia pe |
| 43 | 22 | 75.9 | 77 | 12 | Q8QN71 | Q8qn71 ectocarpus |
| 44 | 22 | 75.9 | 93 | 7 | Q9GJ30 | Q9gj30 homo sapien |
| 45 | 22 | 75.9 | 93 | 7 | Q9GJ32 | Q9gj32 homo sapien |

ALIGNMENTS

RESULT 1
Q92K02
ID Q92K02 PRELIMINARY; PRT; 119 AA.
AC Q92K02;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical transmembrane protein SMc01986.
GN R02520 OR SMC01986.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591791; CAC47099.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 119 AA; 13504 MW; CFFA1042CA34D6A9 CRC64;

Query Match 79.3%; Score 23; DB 16; Length 119;
Best Local Similarity 37.5%; Pred. No. 8e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 HXXXHXXH 8
| | |
Db 11 HAAAHTEH 18

RESULT 2

```
9YE47
D   Q9YE47      PRELIMINARY;      PRT;   152 AA.
C   Q9YE47;
T   01-NOV-1999 (TrEMBLrel. 12, Created)
T   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
T   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
E   Hypothetical protein APE0723.
N   APE0723.
S   Aeropyrum pernix.
C   Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
C   Desulfurococcaceae; Aeropyrum.
X   NCBI_TaxID=56636;
N   [1]
P   SEQUENCE FROM N.A.
C   STRAIN=K1;
X   MEDLINE=99310339; PubMed=10382966;
A   Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
A   Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
A   Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
A   Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
A   Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
A   Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
T   "Complete genome sequence of an aerobic hyper-thermophilic
T   crenarchaeon, Aeropyrum pernix K1.";
L   DNA Res. 6:83-101(1999).
R   EMBL; AP000060; BAA79699.1; -.
W   Hypothetical protein; Complete proteome.
Q   SEQUENCE   152 AA;  16298 MW;  EE300BACCCBA468F CRC64;

 Query Match             79.3%;  Score 23;  DB 17;  Length 152;
 Best Local Similarity   37.5%;  Pred. No. 9.8e+02;
 Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db         15 HSTIHAAH 22


RESULT 3
Q9XWQ9
ID   Q9XWQ9      PRELIMINARY;      PRT;   177 AA.
AC   Q9XWQ9;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Y11D7A.1 protein.
GN   Y11D7A.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Steward C.A.;
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; AL032632; CAA21589.1; -.
DR   WormPep; Y11D7A.1; CE19027.
DR   InterPro; IPR001220; Lectin_legB.
DR   PROSITE; PS00307; LECTIN_LEGUME_BETA; 1.
SQ   SEQUENCE   177 AA;  19422 MW;  349B9DF2D33D17F1 CRC64;

 Query Match             79.3%;  Score 23;  DB 5;  Length 177;
 Best Local Similarity   37.5%;  Pred. No. 1.1e+03;
 Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db        140 HTTVHSSH 147


RESULT 4
Q9WX65
ID   Q9WX65      PRELIMINARY;      PRT;   204 AA.
AC   Q9WX65;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   BGL protein (Fragment).
GN   BGL.
OS   Acetobacter xylinus.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC   Acetobacteraceae; Gluconacetobacter.
OX   NCBI_TaxID=28448;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JCM7664;
RA   Umeda Y., Hirano A., Hon-nami K., Kunito S., Akiyama H., Onizuka T.,
RA   Ikeuchi M., Inoue Y.;
RT   "Conversion of CO2 into cellulose by gene manipulation of microalgae:
RT   cloning of cellulose synthase genes from Acetobacter xylinum.";
RL   (In) Inui T., Anpo M., Izui K., Yanagida S., Yamaguchi T. (eds.);
RL   Advances in chemical conversions for mitigating carbon dioxide,
RL   pp.114:653-656, Elsevier Science, Amsterdam (1998).
DR   EMBL; AB015802; BAA77589.1; -.
DR   InterPro; IPR001764; Glyco_hydro_3N.
DR   Pfam; PF00933; Glyco_hydro_3; 1.
DR   PRINTS; PR00133; GLHYDRLASE3.
FT   NON_TER     204    204
SQ   SEQUENCE   204 AA;  21146 MW;  2CD1050D8E2E720F CRC64;

  Query Match             79.3%;  Score 23;  DB 2;  Length 204;
  Best Local Similarity   37.5%;  Pred. No. 1.2e+03;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db         21 HDAAHAAH 28


RESULT 5
Q9PFU7
ID   Q9PFU7      PRELIMINARY;      PRT;   240 AA.
AC   Q9PFU7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   GMP synthase.
GN   XF0560.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC   Xanthomonadaceae; Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9a5c;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
```

```
A   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
A   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
A   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
A   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
A   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
A   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
A   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
A   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.B.,
A   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
A   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
A   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
A   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
A   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
T   "The genome sequence of the plant pathogen Xylella fastidiosa.";
L   Nature 406:151-159(2000).
R   EMBL; AE003903; AAF83370.1; -.
R   InterPro; IPR000991; GATase_1.
R   Pfam; PF00117; GATase; 1.
R   PRINTS; PR00096; GATASE.
R   PROSITE; PS00442; GATASE_TYPE_I; 1.
W   Complete proteome.
Q   SEQUENCE   240 AA;  26350 MW;  FF81E5EE1EBEEA35 CRC64;

  Query Match             79.3%;  Score 23;  DB 16;  Length 240;
  Best Local Similarity   37.5%;  Pred. No. 1.4e+03;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db        142 HFSAHATH 149


RESULT 6
Q9F502
ID   Q9F502      PRELIMINARY;      PRT;   324 AA.
AC   Q9F502;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Cytochrome c oxidase subunit 2 (EC 1.9.3.1).
GN   COXB2.
OS   Anabaena variabilis.
OC   Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
OX   NCBI_TaxID=1172;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC29413;
RA   Pils D., Schmetterer G.;
RT   "A second cytochrome c oxidase from the cyanobacterium Anabaena sp.
RT   strain ATCC29413, up-regulated under nitrogen fixing conditions.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- COFACTOR: COPPER A (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
DR   EMBL; AJ296086; CAC12660.1; -.
DR   InterPro; IPR001505; Copper_CuA.
DR   InterPro; IPR002429; Cyt_c_ox_2.
DR   Pfam; PF00116; COX2; 1.
DR   Pfam; PF02790; COX2_TM; 1.
DR   PRINTS; PR01166; CYCOXIDASEII.
DR   ProDom; PD000131; Copper_CuA; 1.
DR   PROSITE; PS00078; COX2; 1.
KW   Copper; Oxidoreductase; Transmembrane.
SQ   SEQUENCE   324 AA;  34739 MW;  E16B6CC160899F72 CRC64;

  Query Match             79.3%;  Score 23;  DB 2;  Length 324;
  Best Local Similarity   37.5%;  Pred. No. 1.8e+03;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db        126 HASAHVAH 133


RESULT 7
Q9FD42
ID   Q9FD42      PRELIMINARY;      PRT;   327 AA.
AC   Q9FD42;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Cytochrome c oxidase subunit II.
GN   CTAC.
OS   Anabaena sp. (strain PCC 7120).
OC   Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX   NCBI_TaxID=103690;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=7120;
RA   Jones K.M., Buikema W.J., Haselkorn R.;
RT   "Characterization of a heterocyst-specific cytochrome c oxidase operon
RT   in Anabaena PCC7120.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- COFACTOR: COPPER A (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
DR   EMBL; AF291994; AAG01550.1; -.
DR   InterPro; IPR001505; Copper_CuA.
DR   InterPro; IPR002429; Cyt_c_ox_2.
DR   Pfam; PF00116; COX2; 1.
DR   Pfam; PF02790; COX2_TM; 1.
DR   PRINTS; PR01166; CYCOXIDASEII.
DR   ProDom; PD000131; Copper_CuA; 1.
DR   PROSITE; PS00078; COX2; 1.
KW   Copper; Oxidoreductase; Transmembrane.
SQ   SEQUENCE   327 AA;  35003 MW;  92324730EB7A92F8 CRC64;

  Query Match             79.3%;  Score 23;  DB 2;  Length 327;
  Best Local Similarity   37.5%;  Pred. No. 1.8e+03;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db        126 HASAHVAH 133


RESULT 8
Q8YU45
ID   Q8YU45      PRELIMINARY;      PRT;   327 AA.
AC   Q8YU45;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Cytochrome c oxidase subunit II.
GN   COXB OR ALR2514.
OS   Anabaena sp. (strain PCC 7120).
OC   Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX   NCBI_TaxID=103690;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21595285; PubMed=11759840;
RA   Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA   Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA   Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA   Yasuda M., Tabata S.;
RT   "Complete genomic sequence of the filamentous nitrogen-fixing
RT   cyanobacterium Anabaena sp. strain PCC 7120.";
RL   DNA Res. 8:205-213(2001).
DR   EMBL; AP003589; BAB74213.1; -.
DR   InterPro; IPR001505; Copper_CuA.
DR   InterPro; IPR002429; Cyt_c_ox_2.
DR   Pfam; PF00116; COX2; 1.
DR   Pfam; PF02790; COX2_TM; 1.
DR   PRINTS; PR01166; CYCOXIDASEII.
DR   ProDom; PD000131; Copper_CuA; 1.
```

```
DR   PROSITE; PS00078; COX2; 1.
KW   Complete proteome.
SQ   SEQUENCE   327 AA;  35006 MW;  3369DCBE31BA92E2 CRC64;

  Query Match             79.3%;  Score 23;  DB 16;  Length 327;
  Best Local Similarity   37.5%;  Pred. No. 1.8e+03;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db        126 HASAHVAH 133


RESULT 9
Q8T468
ID   Q8T468      PRELIMINARY;      PRT;   341 AA.
AC   Q8T468;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   AT13703p.
GN   BCDNA:AT13703.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY089327; AAL90065.1; -.
DR   FlyBase; FBgn0063732; BcDNA:AT13703.
SQ   SEQUENCE   341 AA;  34098 MW;  EE6EBF27B2EC618A CRC64;

  Query Match             79.3%;  Score 23;  DB 5;  Length 341;
  Best Local Similarity   37.5%;  Pred. No. 1.9e+03;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db        287 HSHSHAAH 294


RESULT 10
Q8MY62
ID   Q8MY62      PRELIMINARY;      PRT;   341 AA.
AC   Q8MY62;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Casein kinase I.
GN   CS-CKI.
OS   Ciona savignyi.
OC   Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
OC   Phlebobranchia; Cionidae; Ciona.
OX   NCBI_TaxID=51511;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Satou Y., Satoh N.;
RT   "Ciona savignyi maternal gene.";
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB078412; BAC05520.1; -.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   ATP-binding; Kinase; Transferase.
SQ   SEQUENCE   341 AA;  39427 MW;  9BA8B5348512C185 CRC64;

  Query Match             79.3%;  Score 23;  DB 5;  Length 341;
  Best Local Similarity   37.5%;  Pred. No. 1.9e+03;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db        321 HTTQHASH 328


RESULT 11
Q18888
ID   Q18888      PRELIMINARY;      PRT;   342 AA.
AC   Q18888;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical 39.5 kDa protein.
GN   C56C10.10.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Fulton L.;
RT   "The sequence of C. elegans cosmid C56C10.";
RL   Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U29488; AAA68778.1; -.
DR   WormPep; C56C10.10; CE02564.
DR   InterPro; IPR001440; TPR.
DR   Pfam; PF00515; TPR; 2.
KW   Hypothetical protein.
SQ   SEQUENCE   342 AA;  39470 MW;  85EFE2A77F8D766B CRC64;

  Query Match             79.3%;  Score 23;  DB 5;  Length 342;
  Best Local Similarity   37.5%;  Pred. No. 1.9e+03;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db        128 HSHAHTTH 135


RESULT 12
O54173
ID   O54173      PRELIMINARY;      PRT;   382 AA.
AC   O54173;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
```

```
E   Hypothetical protein SCO5944.
N   SCO5944 OR SC7H1.14.
S   Streptomyces coelicolor.
C   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
C   Streptomycineae; Streptomycetaceae; Streptomyces.
X   NCBI_TaxID=1902;
N   [1]
P   SEQUENCE FROM N.A.
C   STRAIN=A3(2) / M145;
X   MEDLINE=21996410; PubMed=12000953;
A   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
A   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
A   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
A   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
A   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
A   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
A   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
A   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
A   Hopwood D.A.;
T   "Complete genome sequence of the model actinomycete Streptomyces
T   coelicolor A3(2).";
L   Nature 417:141-147(2002).
R   EMBL; AL939125; CAA16201.1; -.
W   Hypothetical protein; Complete proteome.
Q   SEQUENCE   382 AA;  41006 MW;  17C54D56069CE871 CRC64;

 Query Match             79.3%;  Score 23;  DB 16;  Length 382;
 Best Local Similarity   37.5%;  Pred. No. 2.1e+03;
 Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

y          1 HXXXHXXH 8
             |   |  |
b        370 HAARHAAH 377


ESULT 13
9KX27
D   Q9KX27      PRELIMINARY;      PRT;   402 AA.
C   Q9KX27;
T   01-OCT-2000 (TrEMBLrel. 15, Created)
T   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
T   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
E   CoxC protein.
N   COXC.
S   Oligotropha carboxidovorans (Pseudomonas carboxydovorans).
G   Plasmid pHCG3.
C   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
C   Bradyrhizobiaceae; Oligotropha.
X   NCBI_TaxID=40137;
N   [1]
P   SEQUENCE FROM N.A.
C   STRAIN=OM5;
A   Moersdorf G.;
L   Submitted (JUL-1993) to the EMBL/GenBank/DDBJ databases.
N   [2]
P   SEQUENCE FROM N.A.
C   STRAIN=OM5;
A   Schuebel U.;
L   Submitted (JUL-1993) to the EMBL/GenBank/DDBJ databases.
N   [3]
P   SEQUENCE FROM N.A.
C   STRAIN=OM5;
X   MEDLINE=95238294; PubMed=7721710;
A   Schuebel U., Kraut M., Moersdorf G., Meyer O.;
T   "Molecular Characterization of the Gene Cluster coxMSL Encoding the
T   Molybdenum-Containing Carbon Monoxide Dehydrogenase of Oligotropha
T   carboxidovorans.";
L   J. Bacteriol. 177:2197-2197(1995).
N   [4]
P   SEQUENCE FROM N.A.
C   STRAIN=OM5;
A   Santiago B., Schuebel U., Egelseer C., Meyer O.;
T   "Sequence analysis, characterization and CO-specific transcription of
T   the cox gene cluster on the megaplasmid pHCG3 of Oligotropha
T   carboxidovorans.";
RL  Gene 236:1157-1247(1999).
RN  [5]
RP  SEQUENCE FROM N.A.
RC  STRAIN=OM5;
RA  Santiago B.;
RL  Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR  EMBL; X82447; CAB76246.1; -.
DR  InterPro; IPR005330; SPNT_Repeat.
DR  Pfam; PF03707; MHYT; 4.
KW  Plasmid.
SQ  SEQUENCE   402 AA;  42864 MW;  4C6108F085CA133D CRC64;

  Query Match             79.3%;  Score 23;  DB 2;  Length 402;
  Best Local Similarity   37.5%;  Pred. No. 2.1e+03;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db        127 HASAHMTH 134


RESULT 14
Q98UI5
ID  Q98UI5      PRELIMINARY;      PRT;   412 AA.
AC  Q98UI5;
DT  01-JUN-2001 (TrEMBLrel. 17, Created)
DT  01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT  01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE  Thyroid transcription factor-1.
GN  LJTTF-1.
OS  Lampetra japonica (Japanese lamprey) (Entosphenus japonicus).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC  Petromyzontiformes; Petromyzontidae; Lethenteron.
OX  NCBI_TaxID=94989;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=Ventral forebrain;
RA  Ogasawara M., Shigetani Y., Suzuki S., Kuratani S., Satoh N.;
RT  "Expression of Thyroid Transcription Factor-1 (TTF-1) Gene in the
RT  Ventral Forebrain and Endostyle of the Agnathan Vertebrate, Lampetra
RT  japonica.";
RL  Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
CC  -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR  EMBL; AB052339; BAB32434.1; -.
DR  HSSP; P23441; 1FTT.
DR  InterPro; IPR001356; Homeobox.
DR  Pfam; PF00046; homeobox; 1.
DR  PRINTS; PR00024; HOMEOBOX.
DR  ProDom; PD000010; Homeobox; 1.
DR  SMART; SM00389; HOX; 1.
DR  PROSITE; PS00027; HOMEOBOX_1; 1.
DR  PROSITE; PS50071; HOMEOBOX_2; 1.
KW  DNA-binding; Homeobox; Nuclear protein.
SQ  SEQUENCE   412 AA;  43509 MW;  EC844185CD89D5EB CRC64;

  Query Match             79.3%;  Score 23;  DB 13;  Length 412;
  Best Local Similarity   37.5%;  Pred. No. 2.2e+03;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 HXXXHXXH 8
              |   |  |
Db        114 HHAAHAAH 121


RESULT 15
Q947K4
ID  Q947K4      PRELIMINARY;      PRT;   466 AA.
AC  Q947K4;
DT  01-DEC-2001 (TrEMBLrel. 19, Created)
DT  01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT  01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
```

E Thiohydroximate S-glucosyltransferase.
S Brassica napus (Rape).
C Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
C eurosids II; Brassicales; Brassicaceae; Brassica.
X NCBI_TaxID=3708;
N [1]
P SEQUENCE FROM N.A.
A Marillia E.-F.A., MacPherson J.M., Tsang E.W.T., Van Audenhove K.,
A Keller W.A., GrootWassink J.W.D.;
T "Molecular cloning of a Brassica napus thiohydroximate S-
T glucosyltransferase gene and its expression in Escherichia coli.";
L Physiol. Plantarum 0:0-0(2001).
R EMBL; AF304430; AAL09350.1; -.
R InterPro; IPR002213; UDP_gluco_trans.
R Pfam; PF00201; UDPGT; 1.
R PROSITE; PS00375; UDPGT; 1.
W Transferase.
Q SEQUENCE 466 AA; 50826 MW; D5991B82129C2C1C CRC64;

Query Match 79.3%; Score 23; DB 10; Length 466;
Best Local Similarity 37.5%; Pred. No. 2.4e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

y 1 HXXXHXXH 8
b 188 HSSSHAEH 195

earch completed: November 21, 2003, 15:49:43
ob time : 29.5 secs

GenCore version 5.1.6


)M protein - protein search, using sw model

tun on: November 21, 2003, 15:28:59 ; Search time 36 Seconds
(without alignments)
35.273 Million cell updates/sec

'itle: US-10-064-903-2
'erfect score: 46
lequence: 1 HDLIHVLH 8

:coring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

learched: 1107863 seqs, 158726573 residues

'otal number of hits satisfying chosen parameters: 1107863

linimum DB seq length: 0
laximum DB seq length: 2000000000

'ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

)atabase : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 42 | 91.3 | 11 | 24 | AAE29884 | Clostridium tetani |
| 2 | 42 | 91.3 | 439 | 23 | ABG69073 | Botulinum neurotox |
| 3 | 42 | 91.3 | 441 | 23 | ABG69068 | Botulinum neurotox |
| 4 | 42 | 91.3 | 441 | 23 | ABG69076 | Botulinum neurotox |
| 5 | 42 | 91.3 | 441 | 23 | ABB80654 | Botulinum toxin ty |
| 6 | 42 | 91.3 | 444 | 23 | ABG69086 | Botulinum neurotox |
| 7 | 42 | 91.3 | 548 | 19 | AAW56014 | Recombinant botuli |
| 8 | 42 | 91.3 | 848 | 23 | ABG69087 | Botulinum neurotox |
| 9 | 42 | 91.3 | 852 | 23 | ABG69077 | Botulinum neurotox |
| 10 | 42 | 91.3 | 858 | 19 | AAW56018 | Recombinant botuli |
| 11 | 42 | 91.3 | 1169 | 19 | AAW56017 | Recombinant botuli |
| 12 | 42 | 91.3 | 1291 | 19 | AAW68392 | Clostridium botuli |
| 13 | 42 | 91.3 | 1315 | 22 | AAB61169 | Clostridium tetani |
| 14 | 38 | 82.6 | 436 | 23 | ABG69072 | Botulinum neurotox |
| 15 | 38 | 82.6 | 443 | 23 | ABG69084 | Botulinum neurotox |
| 16 | 38 | 82.6 | 858 | 23 | ABG69085 | Botulinum neurotox |
| 17 | 38 | 82.6 | 907 | 20 | AAY34888 | Amino acid sequenc |
| 18 | 36 | 78.3 | 104 | 23 | ABP34326 | Human ORF3299 prot |
| 19 | 36 | 78.3 | 422 | 23 | ABG69071 | Botulinum neurotox |
| 20 | 36 | 78.3 | 427 | 23 | ABG69082 | Botulinum neurotox |
| 21 | 36 | 78.3 | 489 | 23 | ABB55296 | Lactococcus lactis |
| 22 | 36 | 78.3 | 773 | 23 | ABB91782 | Herbicidally activ |
| 23 | 36 | 78.3 | 804 | 23 | ABG69083 | Botulinum neurotox |
| 24 | 35 | 76.1 | 60 | 15 | AAR50716 | G-protein coupled |
| 25 | 35 | 76.1 | 60 | 17 | AAW02908 | G-protein coupled |
| 26 | 35 | 76.1 | 155 | 23 | ABU51691 | Helicobacter pylor |
| 27 | 35 | 76.1 | 240 | 23 | ABU51965 | Helicobacter pylor |
| 28 | 35 | 76.1 | 251 | 22 | ABB10271 | Human cDNA SEQ ID |
| 29 | 35 | 76.1 | 251 | 22 | AAU22991 | Novel human enzyme |
| 30 | 35 | 76.1 | 251 | 23 | ABP66858 | Human polypeptide |
| 31 | 35 | 76.1 | 269 | 14 | AAR44805 | Human cyclin D2 ps |
| 32 | 35 | 76.1 | 269 | 24 | ABU03642 | Human expressed pr |
| 33 | 35 | 76.1 | 315 | 18 | AAW20813 | H. pylori secreted |
| 34 | 35 | 76.1 | 737 | 24 | ABU00661 | S. pneumoniae type |
| 35 | 34 | 73.9 | 9 | 22 | AAM98985 | Vaccine related MH |
| 36 | 34 | 73.9 | 118 | 22 | AAB63297 | Human breast cance |
| 37 | 34 | 73.9 | 119 | 23 | ABP34926 | Human ORF3899 prot |
| 38 | 34 | 73.9 | 150 | 22 | AAB63290 | Human breast cance |
| 39 | 34 | 73.9 | 182 | 22 | AAU37971 | Streptococcus pneu |
| 40 | 34 | 73.9 | 198 | 23 | ABB47377 | Listeria monocytog |
| 41 | 34 | 73.9 | 198 | 24 | ABU02143 | S. pneumoniae type |
| 42 | 34 | 73.9 | 226 | 19 | AAY85978 | S. pneumoniae deri |
| 43 | 34 | 73.9 | 452 | 22 | AAU38112 | Salmonella typhi c |
| 44 | 34 | 73.9 | 456 | 22 | AAU34657 | E. coli cellular p |
| 45 | 34 | 73.9 | 498 | 22 | ABG25738 | Novel human diagno |

ALIGNMENTS

RESULT 1
AAE29884
ID AAE29884 standard; peptide; 11 AA.
XX
AC AAE29884;
XX
DT 24-FEB-2003 (first entry)
XX
DE Clostridium tetani TeTx L chain fragment.
XX
KW Gonadotrophin releasing hormone analogue; neurotoxin; prostate cancer;
KW endocrine disorder; gonadotrophin related illness; endometrial cancer;
KW pancreatic cancer; breast cancer; endometriosis; precocious puberty;
KW GnRH-A; therapy; protease; L chain; tetani toxin; TeTx.
XX
OS Clostridium tetani.
XX
PN WO200274327-A2.
XX
PD 26-SEP-2002.
XX
PF 11-MAR-2002; 2002WO-US07379.
XX
PR 15-MAR-2001; 2001US-0810601.
XX
PA (ALLR ) ALLERGAN SALES INC.
XX
PI Donovan S;
XX
DR WPI; 2003-018772/01.
XX
PT New agent comprising a light chain and a (modified) heavy chain of a

```
>T   botulinum, butyricum, or tetani toxin, useful for treating a
>T   gonadotrophin related illness, e.g. breast, prostate pancreatic or
>T   endometrial cancer, or endometriosis  -
<X
>S   Disclosure; Fig 1B; 97pp; English.
<X
:C   The invention relates to an agent comprising a neurotoxin preferably
:C   botulinum toxin for treating endocrine disorders for e.g. gonadotrophin
:C   related illness. The agent is useful for treating gonadotrophin related
:C   illness e.g. prostate cancer, endometrial cancer, pancreatic cancer,
:C   breast cancer, endometriosis or precocious puberty. It is also useful
:C   for decreasing gonadotrophin secretion in a mammal. The present sequence
:C   is Clostridium tetani TeTx (tetani toxin) L-chain fragment. This peptide
:C   is used in the invention.
<X
3Q   Sequence   11 AA;

  Query Match             91.3%;  Score 42;  DB 24;  Length 11;
  Best Local Similarity   87.5%;  Pred. No. 0.27;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||||||
Db          4 HELIHVLH 11


RESULT 2
ABG69073
ID   ABG69073 standard; Protein; 439 AA.
XX
AC   ABG69073;
XX
DT   07-OCT-2002  (first entry)
XX
DE   Botulinum neurotoxin light chain polypeptide #7.
XX
KW   Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW   spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW   bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW   cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW   lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW   immunostimulant; antibacterial.
XX
OS   Clostridium botulinum.
XX
PN   WO200236758-A2.
XX
PD   10-MAY-2002.
XX
PF   06-NOV-2001; 2001WO-US47230.
XX
PR   06-NOV-2000; 2000US-246774P.
PR   20-JUL-2001; 2001US-0910186.
PR   09-AUG-2001; 2001US-311966P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Jensen M;
XX
DR   WPI; 2002-575192/61.
DR   N-PSDB; ABK98543.
XX
PT   Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT   serotype A, useful for producing the neurotoxin for vaccination against
PT   botulism, comprises sequence expressible in host other than Clostridium
PT    -
XX
PS   Claim 33; Page 129-130; 166pp; English.
XX
CC   The invention relates to a nucleic acid molecule encoding a botulinum
CC   neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC   that is expressible in a host organism other than Clostridium, or has a
CC   total A+T content that is less than about 70% The BoNT LC protein is
CC   useful in vaccination against botulism, for eliciting protective immunity
CC   in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC   facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC   myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC   conditions characterised by hyperactivity of the lower motor neuron, and
CC   to control autonomic nerve function or tiptoe-walking due to stiff
CC   muscles common in children with cerebral palsy. The sequences are also
CC   useful for screening for botulinum neurotoxin inhibitors. This sequence
CC   represents a botulinum neurotoxin light chain serotype A protein.
XX
SQ   Sequence   439 AA;

  Query Match             91.3%;  Score 42;  DB 23;  Length 439;
  Best Local Similarity   87.5%;  Pred. No. 13;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||||||
Db        229 HELIHVLH 236


RESULT 3
ABG69068
ID   ABG69068 standard; Protein; 441 AA.
XX
AC   ABG69068;
XX
DT   07-OCT-2002  (first entry)
XX
DE   Botulinum neurotoxin light chain polypeptide #2.
XX
KW   Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW   spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW   bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW   cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW   lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW   immunostimulant; antibacterial.
XX
OS   Clostridium botulinum.
XX
PN   WO200236758-A2.
XX
PD   10-MAY-2002.
XX
PF   06-NOV-2001; 2001WO-US47230.
XX
PR   06-NOV-2000; 2000US-246774P.
PR   20-JUL-2001; 2001US-0910186.
PR   09-AUG-2001; 2001US-311966P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Jensen M;
XX
DR   WPI; 2002-575192/61.
DR   N-PSDB; ABK98538.
XX
PT   Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT   serotype A, useful for producing the neurotoxin for vaccination against
PT   botulism, comprises sequence expressible in host other than Clostridium
PT    -
XX
PS   Claim 33; Page 119-120; 166pp; English.
XX
CC   The invention relates to a nucleic acid molecule encoding a botulinum
CC   neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC   that is expressible in a host organism other than Clostridium, or has a
CC   total A+T content that is less than about 70% The BoNT LC protein is
CC   useful in vaccination against botulism, for eliciting protective immunity
CC   in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC   facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC   myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC   conditions characterised by hyperactivity of the lower motor neuron, and
```

C to control autonomic nerve function or tiptoe-walking due to stiff
C muscles common in children with cerebral palsy. The sequences are also
C useful for screening for botulinum neurotoxin inhibitors. This sequence
C represents a botulinum neurotoxin light chain serotype A protein.
X
Q Sequence 441 AA;

Query Match 91.3%; Score 42; DB 23; Length 441;
Best Local Similarity 87.5%; Pred. No. 13;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Y 1 HDLIHVLH 8
|:||||||
b 230 HELIHVLH 237

ESULT 4
BG69076
D ABG69076 standard; Protein; 441 AA.
X
C ABG69076;
X
T 07-OCT-2002 (first entry)
X
E Botulinum neurotoxin light chain polypeptide #10.
X
W Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
W spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
W bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
W cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
W lower motor neuron hyperactivity; autonomic nerve function; muscular;
W immunostimulant; antibacterial.
X
S Clostridium botulinum.
X
N WO200236758-A2.
X
D 10-MAY-2002.
X
F 06-NOV-2001; 2001WO-US47230.
X
R 06-NOV-2000; 2000US-246774P.
R 20-JUL-2001; 2001US-0910186.
R 09-AUG-2001; 2001US-311966P.
X
A (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
X
I Smith LA, Jensen M;
X
R WPI; 2002-575192/61.
R N-PSDB; ABK98546.
X
T Novel nucleic acid molecule encoding botulinum neurotoxin light chain
T serotype A, useful for producing the neurotoxin for vaccination against
T botulism, comprises sequence expressible in host other than Clostridium
T -
X
S Claim 33; Page 135-136; 166pp; English.
X
C The invention relates to a nucleic acid molecule encoding a botulinum
C neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
C that is expressible in a host organism other than Clostridium, or has a
C total A+T content that is less than about 70% The BoNT LC protein is
C useful in vaccination against botulism, for eliciting protective immunity
C in a mammal, for treating dystonias, spasticity, pain, ocular motility,
C facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
C myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
C conditions characterised by hyperactivity of the lower motor neuron, and
C to control autonomic nerve function or tiptoe-walking due to stiff
C muscles common in children with cerebral palsy. The sequences are also
C useful for screening for botulinum neurotoxin inhibitors. This sequence
C represents a botulinum neurotoxin light chain serotype A protein.
X
SQ Sequence 441 AA;

Query Match 91.3%; Score 42; DB 23; Length 441;
Best Local Similarity 87.5%; Pred. No. 13;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HDLIHVLH 8
|:||||||
Db 230 HELIHVLH 237

RESULT 5
ABB80654
ID ABB80654 standard; peptide; 441 AA.
XX
AC ABB80654;
XX
DT 15-JUL-2002 (first entry)
XX
DE Botulinum toxin type B Danish I light chain.
XX
KW Neurotoxin; biological persistence; dysphonia; strabismus; muscle spasm;
KW dystonia; pain; blepharospasm; hemifacial spasm; excessive salivation;
KW eyelid disorder; cerebral palsy; focal spasticity; spasmodic colitis;
KW neurogenic bladder; anismus; limb spasticity; tic; tremor; bruxism;
KW anal fissure; achalasia; dysphagia; lacrimation; hyperhydrosis; headache;
KW excessive gastrointestinal secretion; botulinum toxin type B; Danish I;
KW light chain.
XX
OS Clostridium botulinum.
XX
PN WO200208268-A2.
XX
PD 31-JAN-2002.
XX
PF 20-JUL-2001; 2001WO-US23122.
XX
PR 21-JUL-2000; 2000US-0620840.
XX
PA (ALLR ) ALLERGAN SALES INC.
XX
PI Steward LE, Fernandez-salas E, Herrington TM, Aoki KR;
XX
DR WPI; 2002-241566/29.
XX
PT Novel modified neurotoxin comprising structural modification which
PT alters the biological persistence and/or biological activity of a
PT neurotoxin, useful for treating neuromuscular or autonomic disorder, or
PT pain -
XX
PS Disclosure; Fig 8; 102pp; English.
XX
CC The sequence represents the botulinum toxin type B Danish I light chain.
CC The invention relates to a novel modified neurotoxin including a
CC structural modification, where the structural modification is effective
CC to alter the biological persistence, or biological activity. The modified
CC neurotoxin is useful for treating spasmodic dysphonia, laryngeal
CC dystonia, oromandibular dysphonia, lingual dystonia, cervical dystonia,
CC focal hand dystonia, blepharospasm, strabismus, hemifacial spasm, eyelid
CC disorder, cerebral palsy, focal spasticity, spasmodic colitis, neurogenic
CC bladder, anismus, limb spasticity, tics, tremors, bruxism, anal fissure,
CC achalasia, dysphagia, lacrimation, hyperhydrosis, excessive salivation,
CC excessive gastrointestinal secretions, pain from muscle spasms, headache
CC pain, brow furrows or skin wrinkles.
XX
SQ Sequence 441 AA;

Query Match 91.3%; Score 42; DB 23; Length 441;
Best Local Similarity 87.5%; Pred. No. 13;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HDLIHVLH 8
|:||||||

```
b       230 HELIHVLH 237


ESULT 6
BG69086
D   ABG69086 standard; Protein; 444 AA.
X
C   ABG69086;
X
T   07-OCT-2002  (first entry)
X
E   Botulinum neurotoxin light chain polypeptide #20.
X
W   Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
W   spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
W   bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
W   cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
W   lower motor neuron hyperactivity; autonomic nerve function; muscular;
W   immunostimulant; antibacterial.
X
S   Clostridium botulinum.
X
N   WO200236758-A2.
X
D   10-MAY-2002.
X
F   06-NOV-2001; 2001WO-US47230.
X
R   06-NOV-2000; 2000US-246774P.
R   20-JUL-2001; 2001US-0910186.
R   09-AUG-2001; 2001US-311966P.
X
A   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
X
I   Smith LA,  Jensen M;
X
R   WPI; 2002-575192/61.
R   N-PSDB; ABK98556.
X
T   Novel nucleic acid molecule encoding botulinum neurotoxin light chain
T   serotype A, useful for producing the neurotoxin for vaccination against
T   botulism, comprises sequence expressible in host other than Clostridium
T    -
X
S   Claim 33; Page 160-161; 166pp; English.
X
C   The invention relates to a nucleic acid molecule encoding a botulinum
C   neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
C   that is expressible in a host organism other than Clostridium, or has a
C   total A+T content that is less than about 70% The BoNT LC protein is
C   useful in vaccination against botulism, for eliciting protective immunity
C   in a mammal, for treating dystonias, spasticity, pain, ocular motility,
C   facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
C   myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
C   conditions characterised by hyperactivity of the lower motor neuron, and
C   to control autonomic nerve function or tiptoe-walking due to stiff
C   muscles common in children with cerebral palsy. The sequences are also
C   useful for screening for botulinum neurotoxin inhibitors. This sequence
C   represents a botulinum neurotoxin light chain serotype A protein.
CX
SQ  Sequence   444 AA;

  Query Match             91.3%;  Score 42;  DB 23;  Length 444;
  Best Local Similarity   87.5%;  Pred. No. 13;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||||||
Db        230 HELIHVLH 237


RESULT 7
AAW56014
ID   AAW56014 standard; Protein; 548 AA.
XX
AC   AAW56014;
XX
DT   27-JUL-1998  (first entry)
XX
DE   Recombinant botulinum neurotoxin type B LH107/B.
XX
KW   Botulinum; recombinant; Clostridium botulinum; neurotoxin;
KW   immunogen; detection; tetanus; non-toxic; toxin.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO9807864-A1.
XX
PD   26-FEB-1998.
XX
PF   22-AUG-1997;   97WO-GB02273.
XX
PR   13-DEC-1996;   96GB-0025996.
PR   23-AUG-1996;   96GB-0017671.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
PA   (SPEY-) SPEYWOOD LAB LTD.
XX
PI   Foster KA,  Quinn CP,  Shone CC;
XX
DR   WPI; 1998-169168/15.
DR   N-PSDB; AAV26286.
XX
PT   Recombinant neurotoxin polypeptides - used to develop therapeutic
PT   agents, immunogens or as non-toxic standards for the detection of
PT   neurotoxins
XX
PS   Example 2; Page 103-104; 137pp; English.
XX
CC   The present sequence represents a recombinant neurotoxin protein from
CC   the present invention. The present invention describes recombinant
CC   neurotoxin proteins which comprise a first and second domain, where
CC   the first domain is adapted to cleave one or more vesicle or
CC   plasma-membrane associated proteins essential to exocytosis, and where
CC   the second domain is adapted: (a) to translocate the protein into a
CC   cell; (b) to increase the solubility of the protein compared to the
CC   solubility of the first domain on its own, or (c) both to translocate
CC   the protein into a cell and to increase the solubility of the protein
CC   compared to the solubility of the first domain on its own, the protein
CC   being free of clostridial neurotoxin (CN) and free of CN precursor that
CC   can be converted into toxin by proteolytic action. The recombinant
CC   proteins can be used as therapeutic agents for targeting cells
CC   expressing a relevant substrate. The products can also be used as
CC   immunogens and as non-toxic standards for the assessment and development
CC   of in vitro assays for the detection of functional botulinum or tetanus
CC   neurotoxins either in foodstuffs or in environmental samples.
XX
SQ   Sequence   548 AA;

  Query Match             91.3%;  Score 42;  DB 19;  Length 548;
  Best Local Similarity   87.5%;  Pred. No. 16;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||||||
Db        230 HELIHVLH 237


RESULT 8
ABG69087
ID   ABG69087 standard; Protein; 848 AA.
XX
AC   ABG69087;
XX
DT   07-OCT-2002  (first entry)
```

```
CX
)E   Botulinum neurotoxin light chain polypeptide #21.
CX
CW   Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
CW   spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
CW   bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
CW   cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
CW   lower motor neuron hyperactivity; autonomic nerve function; muscular;
CW   immunostimulant; antibacterial.
CX
)S   Clostridium botulinum.
CX
)N   WO200236758-A2.
CX
)D   10-MAY-2002.
CX
)F   06-NOV-2001; 2001WO-US47230.
CX
)R   06-NOV-2000; 2000US-246774P.
)R   20-JUL-2001; 2001US-0910186.
)R   09-AUG-2001; 2001US-311966P.
CX
)A   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
CX
)I   Smith LA,  Jensen M;
CX
)R   WPI; 2002-575192/61.
)R   N-PSDB; ABK98557.
CX
)T   Novel nucleic acid molecule encoding botulinum neurotoxin light chain
)T   serotype A, useful for producing the neurotoxin for vaccination against
)T   botulism, comprises sequence expressible in host other than Clostridium
)T    -
CX
)S   Claim 52; Page 162-164; 166pp; English.
CX
)C   The invention relates to a nucleic acid molecule encoding a botulinum
)C   neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
)C   that is expressible in a host organism other than Clostridium, or has a
)C   total A+T content that is less than about 70% The BoNT LC protein is
)C   useful in vaccination against botulism, for eliciting protective immunity
)C   in a mammal, for treating dystonias, spasticity, pain, ocular motility,
)C   facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
)C   myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
)C   conditions characterised by hyperactivity of the lower motor neuron, and
)C   to control autonomic nerve function or tiptoe-walking due to stiff
)C   muscles common in children with cerebral palsy. The sequences are also
)C   useful for screening for botulinum neurotoxin inhibitors. This sequence
)C   represents a botulinum neurotoxin light chain serotype A protein.
CX
)Q   Sequence   848 AA;

  Query Match             91.3%;  Score 42;  DB 23;  Length 848;
  Best Local Similarity   87.5%;  Pred. No. 25;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

)y          1 HDLIHVLH 8
              |:||||||
)b        230 HELIHVLH 237


\ESULT 9
\BG69077
:D   ABG69077 standard; Protein; 852 AA.
CX
\C   ABG69077;
CX
)T   07-OCT-2002  (first entry)
CX
)E   Botulinum neurotoxin light chain polypeptide #11.
CX
CW   Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
CW   spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW   bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW   cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW   lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW   immunostimulant; antibacterial.
XX
OS   Clostridium botulinum.
XX
PN   WO200236758-A2.
XX
PD   10-MAY-2002.
XX
PF   06-NOV-2001; 2001WO-US47230.
XX
PR   06-NOV-2000; 2000US-246774P.
PR   20-JUL-2001; 2001US-0910186.
PR   09-AUG-2001; 2001US-311966P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Jensen M;
XX
DR   WPI; 2002-575192/61.
DR   N-PSDB; ABK98547.
XX
PT   Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT   serotype A, useful for producing the neurotoxin for vaccination against
PT   botulism, comprises sequence expressible in host other than Clostridium
PT    -
XX
PS   Claim 52; Page 138-139; 166pp; English.
XX
CC   The invention relates to a nucleic acid molecule encoding a botulinum
CC   neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC   that is expressible in a host organism other than Clostridium, or has a
CC   total A+T content that is less than about 70% The BoNT LC protein is
CC   useful in vaccination against botulism, for eliciting protective immunity
CC   in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC   facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC   myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC   conditions characterised by hyperactivity of the lower motor neuron, and
CC   to control autonomic nerve function or tiptoe-walking due to stiff
CC   muscles common in children with cerebral palsy. The sequences are also
CC   useful for screening for botulinum neurotoxin inhibitors. This sequence
CC   represents a botulinum neurotoxin light chain serotype A protein.
XX
SQ   Sequence   852 AA;

  Query Match             91.3%;  Score 42;  DB 23;  Length 852;
  Best Local Similarity   87.5%;  Pred. No. 26;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||||||
Db        229 HELIHVLH 236


RESULT 10
AAW56018
ID   AAW56018 standard; Protein; 858 AA.
XX
AC   AAW56018;
XX
DT   27-JUL-1998  (first entry)
XX
DE   Recombinant botulinum neurotoxin type B LH417/B.
XX
KW   Botulinum; recombinant; Clostridium botulinum; neurotoxin;
KW   immunogen; detection; tetanus; non-toxic; toxin.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO9807864-A1.
```

```
XX
PD   26-FEB-1998.
XX
PF   22-AUG-1997;   97WO-GB02273.
XX
PR   13-DEC-1996;   96GB-0025996.
PR   23-AUG-1996;   96GB-0017671.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
PA   (SPEY-) SPEYWOOD LAB LTD.
XX
PI   Foster KA,  Quinn CP,  Shone CC;
XX
DR   WPI; 1998-169168/15.
DR   N-PSDB; AAV26290.
XX
PT   Recombinant neurotoxin polypeptides - used to develop therapeutic
PT   agents, immunogens or as non-toxic standards for the detection of
PT   neurotoxins
XX
PS   Example 2; Page 98-100; 137pp; English.
XX
CC   The present sequence represents a recombinant neurotoxin protein from
CC   the present invention. The present invention describes recombinant
CC   neurotoxin proteins which comprise a first and second domain, where
CC   the first domain is adapted to cleave one or more vesicle or
CC   plasma-membrane associated proteins essential to exocytosis, and where
CC   the second domain is adapted: (a) to translocate the protein into a
CC   cell; (b) to increase the solubility of the protein compared to the
CC   solubility of the first domain on its own, or (c) both to translocate
CC   the protein into a cell and to increase the solubility of the protein
CC   compared to the solubility of the first domain on its own, the protein
CC   being free of clostridial neurotoxin (CN) and free of CN precursor that
CC   can be converted into toxin by proteolytic action. The recombinant
CC   proteins can be used as therapeutic agents for targeting cells
CC   expressing a relevant substrate. The products can also be used as
CC   immunogens and as non-toxic standards for the assessment and development
CC   of in vitro assays for the detection of functional botulinum or tetanus
CC   neurotoxins either in foodstuffs or in environmental samples.
XX
SQ   Sequence   858 AA;

  Query Match             91.3%;  Score 42;  DB 19;  Length 858;
  Best Local Similarity   87.5%;  Pred. No. 26;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||||||
Db        230 HELIHVLH 237


RESULT 11
AAW56017
ID   AAW56017 standard; Protein; 1169 AA.
XX
AC   AAW56017;
XX
DT   27-JUL-1998  (first entry)
XX
DE   Recombinant botulinum neurotoxin type B LH728/B.
XX
KW   Botulinum; recombinant; Clostridium botulinum; neurotoxin;
KW   immunogen; detection; tetanus; non-toxic; toxin.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO9807864-A1.
XX
PD   26-FEB-1998.
XX
PF   22-AUG-1997;   97WO-GB02273.
XX
PR   13-DEC-1996;   96GB-0025996.
PR   23-AUG-1996;   96GB-0017671.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
PA   (SPEY-) SPEYWOOD LAB LTD.
XX
PI   Foster KA,  Quinn CP,  Shone CC;
XX
DR   WPI; 1998-169168/15.
DR   N-PSDB; AAV26289.
XX
PT   Recombinant neurotoxin polypeptides - used to develop therapeutic
PT   agents, immunogens or as non-toxic standards for the detection of
PT   neurotoxins
XX
PS   Example 2; Page 91-94; 137pp; English.
XX
CC   The present sequence represents a recombinant neurotoxin protein from
CC   the present invention. The present invention describes recombinant
CC   neurotoxin proteins which comprise a first and second domain, where
CC   the first domain is adapted to cleave one or more vesicle or
CC   plasma-membrane associated proteins essential to exocytosis, and where
CC   the second domain is adapted: (a) to translocate the protein into a
CC   cell; (b) to increase the solubility of the protein compared to the
CC   solubility of the first domain on its own, or (c) both to translocate
CC   the protein into a cell and to increase the solubility of the protein
CC   compared to the solubility of the first domain on its own, the protein
CC   being free of clostridial neurotoxin (CN) and free of CN precursor that
CC   can be converted into toxin by proteolytic action. The recombinant
CC   proteins can be used as therapeutic agents for targeting cells
CC   expressing a relevant substrate. The products can also be used as
CC   immunogens and as non-toxic standards for the assessment and development
CC   of in vitro assays for the detection of functional botulinum or tetanus
CC   neurotoxins either in foodstuffs or in environmental samples.
XX
SQ   Sequence   1169 AA;

  Query Match             91.3%;  Score 42;  DB 19;  Length 1169;
  Best Local Similarity   87.5%;  Pred. No. 36;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||||||
Db        230 HELIHVLH 237


RESULT 12
AAW68392
ID   AAW68392 standard; Protein; 1291 AA.
XX
AC   AAW68392;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum type B toxin.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW   botulism; BotB.
XX
OS   Clostridium botulinum serotype B Danish strain.
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US15394.
XX
PR   28-AUG-1996;   96US-0704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Thalley BS,  Williams JA;
XX
```

R WPI; 1998-230234/20.
R N-PSDB; AAV30579.
X
T Host cell containing recombinant expression vector encoding
T Clostridium botulinum type B or E toxin - useful to treat humans
T and other animals at risk of intoxication with clostridial toxin
X
S Example 35; Page 291-296; 428pp; English.
X
C This is the amino acid sequence of the type B toxin of Clostridium
C botulinum Danish strain. The C fragment (see AAW68393-94) of the
C serotype B toxin has been expressed as a histidine-tagged protein
C in Escherichia coli. The invention relates to C. botulinum
C recombinant toxins. Methods are provided which allow for the
C isolation of soluble recombinant proteins free of significant
C endotoxin contamination. Preferred hosts for production of the
C recombinant proteins are E. coli, insect cells and yeast cells.
C The recombinant proteins are used as immunogens for the production
C of vaccines and antitoxins that are useful in the treatment of
C humans and animals at risk of intoxication with clostridial toxin.
X
Q Sequence 1291 AA;

Query Match 91.3%; Score 42; DB 19; Length 1291;
Best Local Similarity 87.5%; Pred. No. 39;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HDLIHVLH 8
|:||||||
Db 230 HELIHVLH 237

ESULT 13
AB61169
D AAB61169 standard; Protein; 1315 AA.
X
C AAB61169;
X
T 02-APR-2001 (first entry)
X
E Clostridium tetani TeNT.
X
W Clostridium tetani; TeNT; tetanus toxin; antibacterial; vaccine;
W TeNT fragment C; infection.
X
S Clostridium tetani.
X
N WO200100839-A1.
X
D 04-JAN-2001.
X
F 23-JUN-2000; 2000WO-GB02428.
X
R 25-JUN-1999; 99GB-0014861.
X
A (UNLO ) IMPERIAL COLLEGE SCI TECHNOLOGY & MED.
X
I Fairweather NF, Sinha K;
X
R WPI; 2001-123014/13.
X
T New polypeptide, useful for treating infections of Clostridium tetani,
T comprises tetanus toxin fragment with a mutation in a loop region, -
X
S Disclosure; Page 39; 43pp; English.
X
C The present sequence is given in a specification relating to a novel
C polypeptide comprising tetanus toxin (TeNT) fragment C or its immunogenic
C fragment, containing a mutation in a loop region. The mutation results in
C a reduction in the binding of TeNT fragment C or its immunogenic fragment
C to gangliosides and primary motoneurones, and/or a reduction in the
C ability of TeNT fragment C or its immunogenic fragment to undergo
C retrograde transport. The polypeptide is useful for treating, preventing
CC and reducing the susceptibility to Clostridium tetani infection in a
CC human or animal, and also for producing antibodies which recognise groups
CC within TeNT polypeptides. Antibody produced against the polypeptide is
CC also useful for treating Clostridium tetani infection.
XX
SQ Sequence 1315 AA;

Query Match 91.3%; Score 42; DB 22; Length 1315;
Best Local Similarity 87.5%; Pred. No. 40;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HDLIHVLH 8
|:||||||
Db 233 HELIHVLH 240

RESULT 14
ABG69072
ID ABG69072 standard; Protein; 436 AA.
XX
AC ABG69072;
XX
DT 07-OCT-2002 (first entry)
XX
DE Botulinum neurotoxin light chain polypeptide #6.
XX
KW Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW immunostimulant; antibacterial.
XX
OS Clostridium botulinum.
XX
PN WO200236758-A2.
XX
PD 10-MAY-2002.
XX
PF 06-NOV-2001; 2001WO-US47230.
XX
PR 06-NOV-2000; 2000US-246774P.
PR 20-JUL-2001; 2001US-0910186.
PR 09-AUG-2001; 2001US-311966P.
XX
PA (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Jensen M;
XX
DR WPI; 2002-575192/61.
DR N-PSDB; ABK98542.
XX
PT Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT serotype A, useful for producing the neurotoxin for vaccination against
PT botulism, comprises sequence expressible in host other than Clostridium
PT -
XX
PS Claim 33; Page 127-128; 166pp; English.
XX
CC The invention relates to a nucleic acid molecule encoding a botulinum
CC neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC that is expressible in a host organism other than Clostridium, or has a
CC total A+T content that is less than about 70% The BoNT LC protein is
CC useful in vaccination against botulism, for eliciting protective immunity
CC in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC conditions characterised by hyperactivity of the lower motor neuron, and
CC to control autonomic nerve function or tiptoe-walking due to stiff
CC muscles common in children with cerebral palsy. The sequences are also
CC useful for screening for botulinum neurotoxin inhibitors. This sequence
CC represents a botulinum neurotoxin light chain serotype A protein.
XX

Q Sequence 436 AA;

Query Match 82.6%; Score 38; DB 23; Length 436;
Best Local Similarity 75.0%; Pred. No. 66;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 HDLIHVLH 8
|:||| ||
Db 225 HELIHALH 232

ESULT 15
.BG69084
D ABG69084 standard; Protein; 443 AA.
:X
:C ABG69084;
:X
:T 07-OCT-2002 (first entry)
:X
:E Botulinum neurotoxin light chain polypeptide #18.
:X
:W Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
:W spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
:W bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
:W cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
:W lower motor neuron hyperactivity; autonomic nerve function; muscular;
:W immunostimulant; antibacterial.
:X
:S Clostridium botulinum.
:X
:N WO200236758-A2.
:X
:D 10-MAY-2002.
:X
:F 06-NOV-2001; 2001WO-US47230.
:X
:R 06-NOV-2000; 2000US-246774P.
:R 20-JUL-2001; 2001US-0910186.
:R 09-AUG-2001; 2001US-311966P.
:X
:A (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
:X
:I Smith LA, Jensen M;
:X
:R WPI; 2002-575192/61.
:R N-PSDB; ABK98554.
:X
:T Novel nucleic acid molecule encoding botulinum neurotoxin light chain
:T serotype A, useful for producing the neurotoxin for vaccination against
:T botulism, comprises sequence expressible in host other than Clostridium
:T -
:X
:S Claim 33; Page 155-156; 166pp; English.
:X
:C The invention relates to a nucleic acid molecule encoding a botulinum
:C neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
:C that is expressible in a host organism other than Clostridium, or has a
:C total A+T content that is less than about 70% The BoNT LC protein is
:C useful in vaccination against botulism, for eliciting protective immunity
:C in a mammal, for treating dystonias, spasticity, pain, ocular motility,
:C facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
:C myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
:C conditions characterised by hyperactivity of the lower motor neuron, and
:C to control autonomic nerve function or tiptoe-walking due to stiff
:C muscles common in children with cerebral palsy. The sequences are also
:C useful for screening for botulinum neurotoxin inhibitors. This sequence
:C represents a botulinum neurotoxin light chain serotype A protein.
:X
:Q Sequence 443 AA;

Query Match 82.6%; Score 38; DB 23; Length 443;
Best Local Similarity 75.0%; Pred. No. 67;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 HDLIHVLH 8
|:||| ||
Db 227 HELIHALH 234

Search completed: November 21, 2003, 15:48:02
Job time : 37 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

M protein - protein search, using sw model

un on: November 21, 2003, 15:46:45 ; Search time 14.5 Seconds
(without alignments)
23.344 Million cell updates/sec

itle: US-10-064-903-2
erfect score: 46
equence: 1 HDLIHVLH 8

coring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

earched: 328717 seqs, 42310858 residues

otal number of hits satisfying chosen parameters: 328717

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| esult No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 42 | 91.3 | 548 | 4 | US-09-255-829-24 | Sequence 24, Appl |
| 2 | 42 | 91.3 | 858 | 4 | US-09-255-829-22 | Sequence 22, Appl |
| 3 | 42 | 91.3 | 858 | 4 | US-09-255-829-29 | Sequence 29, Appl |
| 4 | 42 | 91.3 | 1169 | 4 | US-09-255-829-20 | Sequence 20, Appl |
| 5 | 42 | 91.3 | 1315 | 4 | US-08-913-880C-1 | Sequence 1, Appli |
| 6 | 38 | 82.6 | 907 | 4 | US-09-198-452A-306 | Sequence 306, App |
| 7 | 35 | 76.1 | 60 | 1 | US-08-118-270-257 | Sequence 257, App |
| 8 | 35 | 76.1 | 60 | 5 | PCT-US93-08528-257 | Sequence 257, App |
| 9 | 35 | 76.1 | 269 | 5 | PCT-US93-05000-31 | Sequence 31, Appl |
| 10 | 34 | 73.9 | 204 | 4 | US-09-107-532A-7103 | Sequence 7103, Ap |
| 11 | 33 | 71.7 | 60 | 1 | US-08-117-083-20 | Sequence 20, Appl |
| 12 | 33 | 71.7 | 65 | 1 | US-07-879-685B-1 | Sequence 1, Appli |
| 13 | 33 | 71.7 | 162 | 1 | US-07-879-685B-4 | Sequence 4, Appli |
| 14 | 33 | 71.7 | 416 | 1 | US-08-117-083-62 | Sequence 62, Appl |
| 15 | 33 | 71.7 | 431 | 1 | US-08-311-023-2 | Sequence 2, Appli |
| 16 | 33 | 71.7 | 541 | 2 | US-08-540-804-16 | Sequence 16, Appl |
| 17 | 33 | 71.7 | 541 | 2 | US-08-218-265-16 | Sequence 16, Appl |
| 18 | 33 | 71.7 | 541 | 3 | US-08-521-872-16 | Sequence 16, Appl |
| 19 | 33 | 71.7 | 541 | 3 | US-08-590-399-16 | Sequence 16, Appl |
| 20 | 33 | 71.7 | 942 | 4 | US-09-657-931A-9 | Sequence 9, Appli |
| 21 | 32 | 69.6 | 73 | 4 | US-09-328-352-5131 | Sequence 5131, Ap |
| 22 | 32 | 69.6 | 285 | 4 | US-09-328-352-7570 | Sequence 7570, Ap |
| 23 | 32 | 69.6 | 289 | 3 | US-09-105-697-2 | Sequence 2, Appli |
| 24 | 32 | 69.6 | 300 | 4 | US-09-585-858-42 | Sequence 42, Appl |
| 25 | 32 | 69.6 | 310 | 2 | US-08-484-956-88 | Sequence 88, Appl |
| 26 | 32 | 69.6 | 310 | 2 | US-08-757-653-88 | Sequence 88, Appl |
| 27 | 32 | 69.6 | 310 | 4 | US-08-520-946-88 | Sequence 88, Appl |
| 28 | 32 | 69.6 | 315 | 2 | US-08-484-956-91 | Sequence 91, Appl |
| 29 | 32 | 69.6 | 315 | 2 | US-08-757-653-91 | Sequence 91, Appl |
| 30 | 32 | 69.6 | 315 | 4 | US-08-520-946-91 | Sequence 91, Appl |
| 31 | 32 | 69.6 | 320 | 2 | US-08-757-653-163 | Sequence 163, App |
| 32 | 32 | 69.6 | 320 | 2 | US-08-823-516-61 | Sequence 61, Appl |
| 33 | 32 | 69.6 | 320 | 3 | US-08-759-038-102 | Sequence 102, App |
| 34 | 32 | 69.6 | 320 | 3 | US-08-758-314-102 | Sequence 102, App |
| 35 | 32 | 69.6 | 320 | 4 | US-09-684-938-102 | Sequence 102, App |
| 36 | 32 | 69.6 | 320 | 4 | US-09-308-825A-102 | Sequence 102, App |
| 37 | 32 | 69.6 | 322 | 2 | US-08-484-956-89 | Sequence 89, Appl |
| 38 | 32 | 69.6 | 322 | 2 | US-08-757-653-89 | Sequence 89, Appl |
| 39 | 32 | 69.6 | 322 | 4 | US-08-520-946-89 | Sequence 89, Appl |
| 40 | 32 | 69.6 | 359 | 4 | US-09-252-991A-18134 | Sequence 18134, A |
| 41 | 32 | 69.6 | 528 | 2 | US-08-484-956-90 | Sequence 90, Appl |
| 42 | 32 | 69.6 | 528 | 2 | US-08-757-653-90 | Sequence 90, Appl |
| 43 | 32 | 69.6 | 528 | 4 | US-08-520-946-90 | Sequence 90, Appl |
| 44 | 32 | 69.6 | 548 | 2 | US-08-484-956-86 | Sequence 86, Appl |
| 45 | 32 | 69.6 | 548 | 2 | US-08-757-653-86 | Sequence 86, Appl |

ALIGNMENTS

```
RESULT 1
US-09-255-829-24
; Sequence 24, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  548 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-24

  Query Match             91.3%;  Score 42;  DB 4;  Length 548;
  Best Local Similarity   87.5%;  Pred. No. 8.8;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 HDLIHVLH 8
              |:||||||
Db        230 HELIHVLH 237


RESULT 2
US-09-255-829-22
  Sequence 22, Application US/09255829
  Patent No. 6461617
   GENERAL INFORMATION:
     APPLICANT:  Shone, Clifford Charles
     APPLICANT:  Quinn, Conrad Padraig
     APPLICANT:  Foster, Keith Alan
     TITLE OF INVENTION:  Recombinant Toxin Fragments
     NUMBER OF SEQUENCES:  29
     CORRESPONDENCE ADDRESS:
       ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
       STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
       CITY:  WASHINGTON
       STATE:  DC
       COUNTRY:  USA
       ZIP:  20005-3934
     COMPUTER READABLE FORM:
       MEDIUM TYPE:  Floppy disk
       COMPUTER:  IBM PC compatible
       OPERATING SYSTEM:  PC-DOS/MS-DOS
       SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
     CURRENT APPLICATION DATA:
       APPLICATION NUMBER:  US/09/255,829
       FILING DATE:  23-FEB-1999
     PRIOR APPLICATION DATA:
       APPLICATION NUMBER:  PCT/GB97/02273
       FILING DATE:  22-AUG-1997
     PRIOR APPLICATION DATA:
       APPLICATION NUMBER:  US 08/782,893
       FILING DATE:  27-DEC-1996
     ATTORNEY/AGENT INFORMATION:
       NAME:  ESMOND, ROBERT W.
       REGISTRATION NUMBER:  32,893
       REFERENCE/DOCKET NUMBER:  1581.0130002
     TELECOMMUNICATION INFORMATION:
       TELEPHONE:  202-371-2600
       TELEFAX:  202-371-2540
   INFORMATION FOR SEQ ID NO:  22:
     SEQUENCE CHARACTERISTICS:
       LENGTH:  858 amino acids
       TYPE:  amino acid
       TOPOLOGY:  linear
     MOLECULE TYPE:  protein
US-09-255-829-22

  Query Match             91.3%;  Score 42;  DB 4;  Length 858;
  Best Local Similarity   87.5%;  Pred. No. 13;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||||||
Db        230 HELIHVLH 237


RESULT 3
US-09-255-829-29
; Sequence 29, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  858 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-29

  Query Match             91.3%;  Score 42;  DB 4;  Length 858;
  Best Local Similarity   87.5%;  Pred. No. 13;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||||||
Db        230 HELIHVLH 237


RESULT 4
US-09-255-829-20
; Sequence 20, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
```

```
    ATTORNEY/AGENT INFORMATION:
      NAME:  ESMOND, ROBERT W.
      REGISTRATION NUMBER:  32,893
      REFERENCE/DOCKET NUMBER:  1581.0130002
    TELECOMMUNICATION INFORMATION:
      TELEPHONE:  202-371-2600
      TELEFAX:  202-371-2540
  INFORMATION FOR SEQ ID NO:  20:
    SEQUENCE CHARACTERISTICS:
      LENGTH:  1169 amino acids
      TYPE:  amino acid
      TOPOLOGY:  linear
    MOLECULE TYPE:  protein
S-09-255-829-20

 Query Match             91.3%;  Score 42;  DB 4;  Length 1169;
 Best Local Similarity   87.5%;  Pred. No. 18;
 Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

y          1 HDLIHVLH 8
             |:||||||
b        230 HELIHVLH 237


ESULT 5
S-08-913-880C-1
 Sequence 1, Application US/08913880C
 Patent No. 6372225
 GENERAL INFORMATION:
  APPLICANT: MATSUDA, Morihiro
  TITLE OF INVENTION: TETANUS TOXIN FUNCTIONAL FRAGMENT ANTIGEN AND TETANUS
  TITLE OF INVENTION:  VACCINE
  FILE REFERENCE: 216-380P
  CURRENT APPLICATION NUMBER: US/08/913,880C
  CURRENT FILING DATE:  1997-11-24
  NUMBER OF SEQ ID NOS: 17
 SEQ ID NO 1
   LENGTH: 1315
   TYPE: PRT
   ORGANISM: Clostridium tetani
S-08-913-880C-1

 Query Match             91.3%;  Score 42;  DB 4;  Length 1315;
 Best Local Similarity   87.5%;  Pred. No. 20;
 Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

y          1 HDLIHVLH 8
             |:||||||
b        233 HELIHVLH 240


ESULT 6
S-09-198-452A-306
 Sequence 306, Application US/09198452A
 Patent No. 6559294
 GENERAL INFORMATION:
  APPLICANT: Griffais, R.
  TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
  TITLE OF INVENTION:  thereof and uses thereof, in particular for the diagnosis, preve
  TITLE OF INVENTION:  and treatment of infection
  FILE REFERENCE: 9710-003-999
  CURRENT APPLICATION NUMBER: US/09/198,452A
  CURRENT FILING DATE:  1998-11-24
  NUMBER OF SEQ ID NOS: 6849
 SEQ ID NO 306
   LENGTH: 907
   TYPE: PRT
   ORGANISM: Chlamydia pneumoniae
S-09-198-452A-306

 Query Match             82.6%;  Score 38;  DB 4;  Length 907;
 Best Local Similarity   62.5%;  Pred. No. 65;
 Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |||:|:.|
Db        541 HDLLHITH 548


RESULT 7
US-08-118-270-257
; Sequence 257, Application US/08118270
; Patent No. 5508384
; GENERAL INFORMATION:
;   APPLICANT:  Murphy, Randall B.
;   APPLICANT:  Schuster, David I.
;   TITLE OF INVENTION:  POLYPEPTIDES OF G-COUPLED PROTEIN
;   TITLE OF INVENTION:  RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
;   NUMBER OF SEQUENCES:  348
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  BROWDY AND NEIMARK
;     STREET:  419 Seventh Street, N.W., Suite 300
;     CITY:  Washington
;     STATE:  D.C.
;     COUNTRY:  USA
;     ZIP:  20004
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/118,270
;     FILING DATE:  09-SEP-1993
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 07/943,236
;     FILING DATE:  10-SEP-1992
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Townsend, Kevin G.
;     REGISTRATION NUMBER:  34,033
;     REFERENCE/DOCKET NUMBER:  MURPHY=2A
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  202-628-5197
;     TELEFAX:  202-737-3528
;     TELEX:  248633
; INFORMATION FOR SEQ ID NO:  257:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  60 amino acids
;     TYPE:  amino acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  peptide
US-08-118-270-257

  Query Match             76.1%;  Score 35;  DB 1;  Length 60;
  Best Local Similarity   62.5%;  Pred. No. 16;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |||: :||
Db         42 HDLVSILH 49


RESULT 8
PCT-US93-08528-257
; Sequence 257, Application PC/TUS9308528
; GENERAL INFORMATION:
;   APPLICANT:  New York University
;   TITLE OF INVENTION:  POLYPEPTIDES OF G-COUPLED PROTEIN
;   TITLE OF INVENTION:  RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
;   NUMBER OF SEQUENCES:  348
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  BROWDY AND NEIMARK
;     STREET:  419 Seventh Street, N.W., Suite 300
;     CITY:  Washington
;     STATE:  D.C.
```

```
      COUNTRY:  USA
      ZIP:  20004
    COMPUTER READABLE FORM:
      MEDIUM TYPE:  Floppy disk
      COMPUTER:  IBM PC compatible
      OPERATING SYSTEM:  PC-DOS/MS-DOS
      SOFTWARE:  PatentIn Release #1.0, Version #1.25
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER:  PCT/US93/08528
      FILING DATE:  09-SEP-1993
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  US 07/943,236
      FILING DATE:  10-SEP-1992
    ATTORNEY/AGENT INFORMATION:
      NAME:  Townsend, Kevin G.
      REGISTRATION NUMBER:  34,033
      REFERENCE/DOCKET NUMBER:  MURPHY=2 PCT
    TELECOMMUNICATION INFORMATION:
      TELEPHONE:  202-628-5197
      TELEFAX:  202-737-3528
      TELEX:  248633
  INFORMATION FOR SEQ ID NO:  257:
    SEQUENCE CHARACTERISTICS:
      LENGTH:  60 amino acids
      TYPE:  amino acid
      STRANDEDNESS:  single
      TOPOLOGY:  linear
    MOLECULE TYPE:  peptide
CT-US93-08528-257

 Query Match             76.1%;  Score 35;  DB 5;  Length 60;
 Best Local Similarity   62.5%;  Pred. No. 16;
 Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gps    0;

y          1 HDLIHVLH 8
             |||: :||
b         42 HDLVSILH 49


ESULT 9
CT-US93-05000-31
 Sequence 31, Application PC/TUS9305000
  GENERAL INFORMATION:
    APPLICANT:  MITOTIX
    TITLE OF INVENTION:  D-Type Cyclin and Uses Related Thereto
    NUMBER OF SEQUENCES:  42
    CORRESPONDENCE ADDRESS:
      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
      STREET:  Two Militia Drive
      CITY:  Lexington
      STATE:  Massachusetts
      COUNTRY:  US
      ZIP:  02173
    COMPUTER READABLE FORM:
      MEDIUM TYPE:  Floppy disk
      COMPUTER:  IBM PC compatible
      OPERATING SYSTEM:  PC-DOS/MS-DOS
      SOFTWARE:  PatentIn Release #1.0, Version #1.25
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER:  PCT/US93/05000
      FILING DATE:  19930525
      CLASSIFICATION:
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  US/07/888,178
      FILING DATE:  26-MAY-1992
    ATTORNEY/AGENT INFORMATION:
      NAME:  Granahan, Patricia
      REGISTRATION NUMBER:  32,227
      REFERENCE/DOCKET NUMBER:  CSHL91-02A
    TELECOMMUNICATION INFORMATION:
      TELEPHONE:  617-861-6240
      TELEFAX:  616-861-9540
  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  269 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
PCT-US93-05000-31

  Query Match             76.1%;  Score 35;  DB 5;  Length 269;
  Best Local Similarity   62.5%;  Pred. No. 66;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

QY          1 HDLIHVLH 8
              || |::|| 
Db        158 HDFIYILH 165


RESULT 10
US-09-107-532A-7103
; Sequence 7103, Application US/09107532A
; Patent No. 6583275
;   GENERAL INFORMATION:
;        APPLICANT: Lynn A Doucette-Stamm  and David Bush
;        TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;                            ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
;        NUMBER OF SEQUENCES: 7310
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;             STREET: 100 Beaver Street
;             CITY: Waltham
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02354
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: CD/ROM ISO9660
;             COMPUTER: PC
;             OPERATING SYSTEM: <Unknown>
;             SOFTWARE: ASCII
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/107,532A
;             FILING DATE: 30-Jun-1998
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/085,598
;             FILING DATE: 14 May 1998
;             APPLICATION NUMBER: 60/051571
;             FILING DATE: July 2, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ariniello, Pamela Deneke
;             REGISTRATION NUMBER: 40,489
;             REFERENCE/DOCKET NUMBER: GTC-012
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (781)893-5007
;             TELEFAX: (781)893-8277
;   INFORMATION FOR SEQ ID NO: 7103:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 204 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        HYPOTHETICAL: YES
;        ORIGINAL SOURCE:
;             ORGANISM: Enterococcus faecium
;        FEATURE:
;             NAME/KEY:  misc_feature
;             LOCATION: (B) LOCATION 1...204
;        SEQUENCE DESCRIPTION: SEQ ID NO: 7103:
US-09-107-532A-7103

  Query Match             73.9%;  Score 34;  DB 4;  Length 204;
  Best Local Similarity   75.0%;  Pred. No. 74;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

QY          1 HDLIHVLH 8
              | | ||||
```

```
)b        107 HGLYHVLH 114


:ESULT 11
/S-08-117-083-20
  Sequence 20, Application US/08117083
  Patent No. 5719054
   GENERAL INFORMATION:
     APPLICANT:  Boursnell, Michael E.
     APPLICANT:  Inglis, Stephen C.
     APPLICANT:  Munro, Alan J.
     TITLE OF INVENTION:  Recombinant Virus Vectors Encoding Human
     TITLE OF INVENTION:  Papilloma Virus Proteins
     NUMBER OF SEQUENCES:  70
     CORRESPONDENCE ADDRESS:
       ADDRESSEE:  Walter H. Dreger
       STREET:  4 Embarcadero Center, Suite 3400
       CITY:  San Francisco
       STATE:  CA
       COUNTRY:  USA
       ZIP:  94111
     COMPUTER READABLE FORM:
       MEDIUM TYPE:  Floppy disk
       COMPUTER:  IBM PC compatible
       OPERATING SYSTEM:  PC-DOS/MS-DOS
       SOFTWARE:  PatentIn Release #1.0, Version #1.25
     CURRENT APPLICATION DATA:
       APPLICATION NUMBER:  US/08/117,083
       FILING DATE:  10-SEP-1993
       CLASSIFICATION:  435
     ATTORNEY/AGENT INFORMATION:
       NAME:  Dreger, Walter H.
       REGISTRATION NUMBER:  24,190
       REFERENCE/DOCKET NUMBER:  A-58783
     TELECOMMUNICATION INFORMATION:
       TELEPHONE:  415-781-1989
       TELEFAX:  415-398-3249
       TELEX:  910 277299
    INFORMATION FOR SEQ ID NO:  20:
     SEQUENCE CHARACTERISTICS:
       LENGTH:  60 amino acids
       TYPE:  amino acid
       STRANDEDNESS:  single
       TOPOLOGY:  linear
     MOLECULE TYPE:  protein
     FEATURE:
       NAME/KEY:  Protein
       LOCATION:  1..60
       OTHER INFORMATION:  /note= "Xaa refers to stop codon in
       OTHER INFORMATION:  the open reading frame."
/S-08-117-083-20

  Query Match             71.7%;  Score 33;  DB 1;  Length 60;
  Best Local Similarity   85.7%;  Pred. No. 34;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

)y          2 DLIHVLH 8
              |||| ||
)b         24 DLIHFLH 30


:ESULT 12
/S-07-879-685B-1
  Sequence 1, Application US/07879685B
  Patent No. 5296383
   GENERAL INFORMATION:
     APPLICANT:  DAIKIN INDUSTRIES, LTD.
     TITLE OF INVENTION:  A human centromere antigen
     TITLE OF INVENTION:  polypeptide
     NUMBER OF SEQUENCES:  4
     CORRESPONDENCE ADDRESS:
       ADDRESSEE:  Umeda Center Building, 4-12
       STREET:  Nakazaki-nishi, 2-chome
       CITY:  Kita-ku
       STATE:  Osaka
       COUNTRY:  Japan
       ZIP:  530
     COMPUTER READABLE FORM:
       MEDIUM TYPE:  Floppy disk
       COMPUTER:  IBM PC compatible
       OPERATING SYSTEM:  PC-DOS/MS-DOS
       SOFTWARE:  PatentIn Release #1.0, Version #1.25
     CURRENT APPLICATION DATA:
       APPLICATION NUMBER:  US/07/879,685B
       FILING DATE:  19920507
       CLASSIFICATION:  436
     PRIOR APPLICATION DATA:
       APPLICATION NUMBER:  JP 3-102517
       FILING DATE:  08-May-1991
    INFORMATION FOR SEQ ID NO:  1:
     SEQUENCE CHARACTERISTICS:
       LENGTH:  65 amino acids
       TYPE:  AMINO ACID
       TOPOLOGY:  linear
     MOLECULE TYPE:  peptide
     FRAGMENT TYPE:  internal fragment
     ORIGINAL SOURCE:
       ORGANISM:  human
US-07-879-685B-1

  Query Match             71.7%;  Score 33;  DB 1;  Length 65;
  Best Local Similarity   83.3%;  Pred. No. 37;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HDLIHV 6
                |||:||
Db         42 HDLVHV 47


RESULT 13
US-07-879-685B-4
; Sequence 4, Application US/07879685B
; Patent No. 5296383
;  GENERAL INFORMATION:
;    APPLICANT:  DAIKIN INDUSTRIES, LTD.
;    TITLE OF INVENTION:  A human centromere antigen
;    TITLE OF INVENTION:  polypeptide
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Umeda Center Building, 4-12
;      STREET:  Nakazaki-nishi, 2-chome
;      CITY:  Kita-ku
;      STATE:  Osaka
;      COUNTRY:  Japan
;      ZIP:  530
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/879,685B
;      FILING DATE:  19920507
;      CLASSIFICATION:  436
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 3-102517
;      FILING DATE:  08-May-1991
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  162 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-879-685B-4

  Query Match             71.7%;  Score 33;  DB 1;  Length 162;
```

```
  Best Local Similarity   83.3%;  Pred. No. 87;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HDLIHV 6
              |||:||
Db        139 HDLVHV 144


RESULT 14
US-08-117-083-62
 Sequence 62, Application US/08117083
 Patent No. 5719054
  GENERAL INFORMATION:
    APPLICANT:  Boursnell, Michael E.
    APPLICANT:  Inglis, Stephen C.
    APPLICANT:  Munro, Alan J.
    TITLE OF INVENTION:  Recombinant Virus Vectors Encoding Human
    TITLE OF INVENTION:  Papilloma Virus Proteins
    NUMBER OF SEQUENCES:  70
    CORRESPONDENCE ADDRESS:
      ADDRESSEE:  Walter H. Dreger
      STREET:  4 Embarcadero Center, Suite 3400
      CITY:  San Francisco
      STATE:  CA
      COUNTRY:  USA
      ZIP:  94111
    COMPUTER READABLE FORM:
      MEDIUM TYPE:  Floppy disk
      COMPUTER:  IBM PC compatible
      OPERATING SYSTEM:  PC-DOS/MS-DOS
      SOFTWARE:  PatentIn Release #1.0, Version #1.25
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER:  US/08/117,083
      FILING DATE:  10-SEP-1993
      CLASSIFICATION:  435
    ATTORNEY/AGENT INFORMATION:
      NAME:  Dreger, Walter H.
      REGISTRATION NUMBER:  24,190
      REFERENCE/DOCKET NUMBER:  A-58783
    TELECOMMUNICATION INFORMATION:
      TELEPHONE:  415-781-1989
      TELEFAX:  415-398-3249
      TELEX:  910 277299
  INFORMATION FOR SEQ ID NO:  62:
    SEQUENCE CHARACTERISTICS:
      LENGTH:  416 amino acids
      TYPE:  amino acid
      STRANDEDNESS:  single
      TOPOLOGY:  linear
    MOLECULE TYPE:  protein
    FEATURE:
      NAME/KEY:  Protein
      LOCATION:  1..416
      OTHER INFORMATION:  /note= "Xaa refers to stop codon in
      OTHER INFORMATION:  the open reading frame."
US-08-117-083-62

  Query Match             71.7%;  Score 33;  DB 1;  Length 416;
  Best Local Similarity   85.7%;  Pred. No. 2.1e+02;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 DLIHVLH 8
              |||| ||
Db        307 DLIHFLH 313


RESULT 15
US-08-311-023-2
; Sequence 2, Application US/08311023
; Patent No. 5693465
;  GENERAL INFORMATION:
;    APPLICANT:  MANNING, David Lockwood
;    APPLICANT:  NICHOLSON, Robert Ian
;    APPLICANT:  GEE, Julia Margaret
;    APPLICANT:  GREEN, Christopher Douglas
;    TITLE OF INVENTION:  METHODS FOR PREDICTING THE BEHAVIOUR OF
;    TITLE OF INVENTION:  BREAST TUMOURS
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Young & Thompson
;      STREET:  745 South 23rd Street
;      CITY:  Arlington
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/311,023
;      FILING DATE:  22-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PATCH, Andrew J.
;      REGISTRATION NUMBER:  Reg. No. 5693465 32,925
;      REFERENCE/DOCKET NUMBER:  WCM.56
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703/521-2297
;      TELEFAX:  703/685-0573
;      TELEX:  248425
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  431 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-311-023-2

  Query Match             71.7%;  Score 33;  DB 1;  Length 431;
  Best Local Similarity   62.5%;  Pred. No. 2.2e+02;
  Matches    5;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              ||  |:||
Db        232 HDYHHILH 239


Search completed: November 21, 2003, 15:51:04
Job time : 15.5 secs
```

GenCore version 5.1.6


OM protein - protein search, using sw model

Run on: November 21, 2003, 15:49:51 ; Search time 23.5 Seconds
(without alignments)
62.148 Million cell updates/sec

Title: US-10-064-903-2
Perfect score: 46
Sequence: 1 HDLIHVLH 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 46 | 100.0 | 8 | 15 | US-10-064-903-2 | Sequence 2, Appli |
| 2 | 42 | 91.3 | 439 | 14 | US-10-011-588-17 | Sequence 17, Appl |
| 3 | 42 | 91.3 | 441 | 11 | US-09-910-346C-20 | Sequence 20, Appl |
| 4 | 42 | 91.3 | 441 | 14 | US-10-011-588-7 | Sequence 7, Appli |
| 5 | 42 | 91.3 | 441 | 14 | US-10-011-588-23 | Sequence 23, Appl |
| 6 | 42 | 91.3 | 444 | 14 | US-10-011-588-43 | Sequence 43, Appl |
| 7 | 42 | 91.3 | 458 | 12 | US-10-241-596-114 | Sequence 114, App |
| 8 | 42 | 91.3 | 548 | 12 | US-10-241-596-24 | Sequence 24, Appl |
| 9 | 42 | 91.3 | 848 | 14 | US-10-011-588-45 | Sequence 45, Appl |
| 10 | 42 | 91.3 | 852 | 14 | US-10-011-588-25 | Sequence 25, Appl |
| 11 | 42 | 91.3 | 858 | 12 | US-10-241-596-22 | Sequence 22, Appl |
| 12 | 42 | 91.3 | 860 | 12 | US-10-241-596-175 | Sequence 175, App |
| 13 | 42 | 91.3 | 862 | 12 | US-10-241-596-94 | Sequence 94, Appl |
| 14 | 42 | 91.3 | 862 | 12 | US-10-241-596-171 | Sequence 171, App |
| 15 | 42 | 91.3 | 862 | 12 | US-10-241-596-173 | Sequence 173, App |
| 16 | 42 | 91.3 | 864 | 12 | US-10-241-596-102 | Sequence 102, App |
| 17 | 42 | 91.3 | 865 | 12 | US-10-241-596-100 | Sequence 100, App |
| 18 | 42 | 91.3 | 866 | 12 | US-10-241-596-88 | Sequence 88, Appl |
| 19 | 42 | 91.3 | 866 | 12 | US-10-241-596-104 | Sequence 104, App |
| 20 | 42 | 91.3 | 867 | 12 | US-10-241-596-80 | Sequence 80, Appl |
| 21 | 42 | 91.3 | 867 | 12 | US-10-241-596-96 | Sequence 96, Appl |
| 22 | 42 | 91.3 | 867 | 12 | US-10-241-596-98 | Sequence 98, Appl |
| 23 | 42 | 91.3 | 870 | 12 | US-10-241-596-92 | Sequence 92, Appl |
| 24 | 42 | 91.3 | 871 | 12 | US-10-241-596-84 | Sequence 84, Appl |
| 25 | 42 | 91.3 | 871 | 12 | US-10-241-596-86 | Sequence 86, Appl |
| 26 | 42 | 91.3 | 871 | 12 | US-10-241-596-90 | Sequence 90, Appl |
| 27 | 42 | 91.3 | 872 | 12 | US-10-241-596-145 | Sequence 145, App |
| 28 | 42 | 91.3 | 876 | 12 | US-10-241-596-82 | Sequence 82, Appl |
| 29 | 42 | 91.3 | 876 | 12 | US-10-241-596-106 | Sequence 106, App |
| 30 | 42 | 91.3 | 876 | 12 | US-10-241-596-108 | Sequence 108, App |
| 31 | 42 | 91.3 | 879 | 12 | US-10-241-596-143 | Sequence 143, App |
| 32 | 42 | 91.3 | 887 | 12 | US-10-241-596-147 | Sequence 147, App |
| 33 | 42 | 91.3 | 888 | 12 | US-10-241-596-112 | Sequence 112, App |
| 34 | 42 | 91.3 | 1169 | 12 | US-10-241-596-20 | Sequence 20, Appl |
| 35 | 42 | 91.3 | 1315 | 12 | US-10-241-596-141 | Sequence 141, App |
| 36 | 42 | 91.3 | 1420 | 12 | US-10-241-596-110 | Sequence 110, App |
| 37 | 38 | 82.6 | 436 | 14 | US-10-011-588-15 | Sequence 15, Appl |
| 38 | 38 | 82.6 | 443 | 14 | US-10-011-588-39 | Sequence 39, Appl |
| 39 | 38 | 82.6 | 858 | 14 | US-10-011-588-41 | Sequence 41, Appl |
| 40 | 36 | 78.3 | 422 | 14 | US-10-011-588-13 | Sequence 13, Appl |
| 41 | 36 | 78.3 | 427 | 14 | US-10-011-588-35 | Sequence 35, Appl |
| 42 | 36 | 78.3 | 804 | 14 | US-10-011-588-37 | Sequence 37, Appl |
| 43 | 35 | 76.1 | 251 | 9 | US-09-764-853-579 | Sequence 579, App |
| 44 | 35 | 76.1 | 302 | 12 | US-10-259-165-72 | Sequence 72, Appl |
| 45 | 35 | 76.1 | 302 | 12 | US-10-259-165-410 | Sequence 410, App |

ALIGNMENTS

RESULT 1
US-10-064-903-2
; Sequence 2, Application US/10064903
; Publication No. US20030059912A1
; GENERAL INFORMATION:
; APPLICANT: Biotecon Gesellschaft fur biotechnologische Entwicklung und Consulting
; APPLICANT: mbH
; TITLE OF INVENTION: HYBRID PROTEIN FOR INHIBITING THE DEGRANULATION OF MASTOCYTES AND
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: BIO-001PCT-CIP
; CURRENT APPLICATION NUMBER: US/10/064,903
; CURRENT FILING DATE: 2002-08-27
; PRIOR APPLICATION NUMBER: US 09/700,540
; PRIOR FILING DATE: 2001-01-19
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Clostridium tetani
US-10-064-903-2

Query Match 100.0%; Score 46; DB 15; Length 8;
Best Local Similarity 100.0%; Pred. No. 5.9e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HDLIHVLH 8
||||||||
Db 1 HDLIHVLH 8

RESULT 2
US-10-011-588-17
; Sequence 17, Application US/10011588
; Publication No. US20020168727A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Leonard
; APPLICANT: Jensen, Melody

```
;  TITLE OF INVENTION: RECOMBINANT LIGHT CHAINS OF BOTULINUM
;  TITLE OF INVENTION:  NEUROTOXINS AND LIGHT CHAIN FUSION PROTEINS FOR USE IN
;  TITLE OF INVENTION:  RESEARCH AND CLINICAL THERAPY
;  FILE REFERENCE: A34796 067252.0113
;  CURRENT APPLICATION NUMBER: US/10/011,588
;  CURRENT FILING DATE:  2002-03-29
;  PRIOR APPLICATION NUMBER: 09/910,186
;  PRIOR FILING DATE: 2001-07-20
;  PRIOR APPLICATION NUMBER: 09/611,419
;  PRIOR FILING DATE: 2000-07-06
;  PRIOR APPLICATION NUMBER: 60/246,744
;  PRIOR FILING DATE: 2000-11-06
;  PRIOR APPLICATION NUMBER: 60/311,966
;  PRIOR FILING DATE: 2001-08-09
;  NUMBER OF SEQ ID NOS: 47
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 439
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic botulinum neurotoxin light chain of
;   OTHER INFORMATION: serotype G based on wild-type Clostridium
;   OTHER INFORMATION: botulinum sequence
JS-10-011-588-17

  Query Match             91.3%;  Score 42;  DB 14;  Length 439;
  Best Local Similarity   87.5%;  Pred. No. 14;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||||||
Db        229 HELIHVLH 236


RESULT 3
JS-09-910-346C-20
; Sequence 20, Application US/09910346C
; Publication No. US20030027752A1
; GENERAL INFORMATION:
;  APPLICANT: STEWARD, LANCE E
;  APPLICANT:  FERNANDEZ-SALAS, ESTER
;  APPLICANT:  HERRINGTON, TODD M
;  APPLICANT:  AOKI, KEI R
;  TITLE OF INVENTION: Leucine-based motif and clostridial neurotoxins
;  FILE REFERENCE: D-2885CIP
;  CURRENT APPLICATION NUMBER: US/09/910,346C
;  CURRENT FILING DATE:  2000-07-21
;  PRIOR APPLICATION NUMBER: US 09/620,840
;  PRIOR FILING DATE: 2000-07-21
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
;   LENGTH: 441
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-09-910-346C-20

  Query Match             91.3%;  Score 42;  DB 11;  Length 441;
  Best Local Similarity   87.5%;  Pred. No. 14;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||||||
Db        230 HELIHVLH 237


RESULT 4
US-10-011-588-7
; Sequence 7, Application US/10011588
; Publication No. US20020168727A1
; GENERAL INFORMATION:
;  APPLICANT: Smith, Leonard
;  APPLICANT:  Jensen, Melody
;  TITLE OF INVENTION: RECOMBINANT LIGHT CHAINS OF BOTULINUM
;  TITLE OF INVENTION:  NEUROTOXINS AND LIGHT CHAIN FUSION PROTEINS FOR USE IN
;  TITLE OF INVENTION:  RESEARCH AND CLINICAL THERAPY
;  FILE REFERENCE: A34796 067252.0113
;  CURRENT APPLICATION NUMBER: US/10/011,588
;  CURRENT FILING DATE:  2002-03-29
;  PRIOR APPLICATION NUMBER: 09/910,186
;  PRIOR FILING DATE: 2001-07-20
;  PRIOR APPLICATION NUMBER: 09/611,419
;  PRIOR FILING DATE: 2000-07-06
;  PRIOR APPLICATION NUMBER: 60/246,744
;  PRIOR FILING DATE: 2000-11-06
;  PRIOR APPLICATION NUMBER: 60/311,966
;  PRIOR FILING DATE: 2001-08-09
;  NUMBER OF SEQ ID NOS: 47
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
;   LENGTH: 441
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic botulinum neurotoxin light chain of
;   OTHER INFORMATION: serotype B based on wild-type Clostridium
;   OTHER INFORMATION: botulinum sequence
US-10-011-588-7

  Query Match             91.3%;  Score 42;  DB 14;  Length 441;
  Best Local Similarity   87.5%;  Pred. No. 14;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||||||
Db        230 HELIHVLH 237


RESULT 5
US-10-011-588-23
; Sequence 23, Application US/10011588
; Publication No. US20020168727A1
; GENERAL INFORMATION:
;  APPLICANT: Smith, Leonard
;  APPLICANT:  Jensen, Melody
;  TITLE OF INVENTION: RECOMBINANT LIGHT CHAINS OF BOTULINUM
;  TITLE OF INVENTION:  NEUROTOXINS AND LIGHT CHAIN FUSION PROTEINS FOR USE IN
;  TITLE OF INVENTION:  RESEARCH AND CLINICAL THERAPY
;  FILE REFERENCE: A34796 067252.0113
;  CURRENT APPLICATION NUMBER: US/10/011,588
;  CURRENT FILING DATE:  2002-03-29
;  PRIOR APPLICATION NUMBER: 09/910,186
;  PRIOR FILING DATE: 2001-07-20
;  PRIOR APPLICATION NUMBER: 09/611,419
;  PRIOR FILING DATE: 2000-07-06
;  PRIOR APPLICATION NUMBER: 60/246,744
;  PRIOR FILING DATE: 2000-11-06
;  PRIOR APPLICATION NUMBER: 60/311,966
;  PRIOR FILING DATE: 2001-08-09
;  NUMBER OF SEQ ID NOS: 47
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
;   LENGTH: 441
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant protein encoded by SEQ ID NO:22
US-10-011-588-23

  Query Match             91.3%;  Score 42;  DB 14;  Length 441;
  Best Local Similarity   87.5%;  Pred. No. 14;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||||||
```

)b 230 HELIHVLH 237

:ESULT 6
)S-10-011-588-43
Sequence 43, Application US/10011588
Publication No. US20020168727A1
GENERAL INFORMATION:
APPLICANT: Smith, Leonard
APPLICANT: Jensen, Melody
TITLE OF INVENTION: RECOMBINANT LIGHT CHAINS OF BOTULINUM
TITLE OF INVENTION: NEUROTOXINS AND LIGHT CHAIN FUSION PROTEINS FOR USE IN
TITLE OF INVENTION: RESEARCH AND CLINICAL THERAPY
FILE REFERENCE: A34796 067252.0113
CURRENT APPLICATION NUMBER: US/10/011,588
CURRENT FILING DATE: 2002-03-29
PRIOR APPLICATION NUMBER: 09/910,186
PRIOR FILING DATE: 2001-07-20
PRIOR APPLICATION NUMBER: 09/611,419
PRIOR FILING DATE: 2000-07-06
PRIOR APPLICATION NUMBER: 60/246,744
PRIOR FILING DATE: 2000-11-06
PRIOR APPLICATION NUMBER: 60/311,966
PRIOR FILING DATE: 2001-08-09
NUMBER OF SEQ ID NOS: 47
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 43
LENGTH: 444
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant protein encoded by SEQ ID NO:42
NAME/KEY: UNSURE
LOCATION: (442)...(443)
OTHER INFORMATION: Any amino acid at each position
)S-10-011-588-43

Query Match 91.3%; Score 42; DB 14; Length 444;
Best Local Similarity 87.5%; Pred. No. 14;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

)y 1 HDLIHVLH 8
|:||||||
)b 230 HELIHVLH 237

:ESULT 7
)S-10-241-596-114
Sequence 114, Application US/10241596
Publication No. US20030166238A1
GENERAL INFORMATION:
APPLICANT: Microbiological Research Authority
APPLICANT: The Speywood Laboratory Limited
TITLE OF INVENTION: Recombinant Toxin Fragments
FILE REFERENCE: 1581.0130003
CURRENT APPLICATION NUMBER: US/10/241,596
CURRENT FILING DATE: 2002-09-12
PRIOR APPLICATION NUMBER: US 09/255,829
PRIOR FILING DATE: 1999-02-23
PRIOR APPLICATION NUMBER: US 09/242,689
PRIOR FILING DATE: 1999-02-23
PRIOR APPLICATION NUMBER: PCT/GB97/02273
PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 08/782,893
PRIOR FILING DATE: 1996-12-27
PRIOR APPLICATION NUMBER: GB 9625996.5
PRIOR FILING DATE: 1996-12-13
PRIOR APPLICATION NUMBER: GB 9617671.4
PRIOR FILING DATE: 1996-08-23
NUMBER OF SEQ ID NOS: 175
SOFTWARE: PatentIn version 3.1
SEQ ID NO 114
LENGTH: 458
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-241-596-114

Query Match 91.3%; Score 42; DB 12; Length 458;
Best Local Similarity 87.5%; Pred. No. 15;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HDLIHVLH 8
|:||||||
Db 232 HELIHVLH 239

RESULT 8
US-10-241-596-24
; Sequence 24, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
; APPLICANT: Microbiological Research Authority
; APPLICANT: The Speywood Laboratory Limited
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130003
; CURRENT APPLICATION NUMBER: US/10/241,596
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 09/255,829
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: US 09/242,689
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/GB97/02273
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 08/782,893
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB 9625996.5
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: GB 9617671.4
; PRIOR FILING DATE: 1996-08-23
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
; LENGTH: 548
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-241-596-24

Query Match 91.3%; Score 42; DB 12; Length 548;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HDLIHVLH 8
|:||||||
Db 230 HELIHVLH 237

RESULT 9
US-10-011-588-45
; Sequence 45, Application US/10011588
; Publication No. US20020168727A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Leonard
; APPLICANT: Jensen, Melody
; TITLE OF INVENTION: RECOMBINANT LIGHT CHAINS OF BOTULINUM
; TITLE OF INVENTION: NEUROTOXINS AND LIGHT CHAIN FUSION PROTEINS FOR USE IN
; TITLE OF INVENTION: RESEARCH AND CLINICAL THERAPY
; FILE REFERENCE: A34796 067252.0113
; CURRENT APPLICATION NUMBER: US/10/011,588
; CURRENT FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: 09/910,186
; PRIOR FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: 09/611,419
; PRIOR FILING DATE: 2000-07-06
; PRIOR APPLICATION NUMBER: 60/246,744
; PRIOR FILING DATE: 2000-11-06
; PRIOR APPLICATION NUMBER: 60/311,966

```
  PRIOR FILING DATE: 2001-08-09
  NUMBER OF SEQ ID NOS: 47
  SOFTWARE: FastSEQ for Windows Version 4.0
 SEQ ID NO 45
   LENGTH: 848
   TYPE: PRT
   ORGANISM: Artificial Sequence
   FEATURE:
   OTHER INFORMATION: Recombinant protein encoded by SEQ ID NO:44
US-10-011-588-45

  Query Match             91.3%;  Score 42;  DB 14;  Length 848;
  Best Local Similarity   87.5%;  Pred. No. 27;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||||||
Db        230 HELIHVLH 237


RESULT 10
US-10-011-588-25
; Sequence 25, Application US/10011588
; Publication No. US20020168727A1
; GENERAL INFORMATION:
;  APPLICANT: Smith, Leonard
;  APPLICANT:  Jensen, Melody
;  TITLE OF INVENTION: RECOMBINANT LIGHT CHAINS OF BOTULINUM
;  TITLE OF INVENTION:  NEUROTOXINS AND LIGHT CHAIN FUSION PROTEINS FOR USE IN
;  TITLE OF INVENTION:  RESEARCH AND CLINICAL THERAPY
;  FILE REFERENCE: A34796 067252.0113
;  CURRENT APPLICATION NUMBER: US/10/011,588
;  CURRENT FILING DATE:  2002-03-29
;  PRIOR APPLICATION NUMBER: 09/910,186
;  PRIOR FILING DATE: 2001-07-20
;  PRIOR APPLICATION NUMBER: 09/611,419
;  PRIOR FILING DATE: 2000-07-06
;  PRIOR APPLICATION NUMBER: 60/246,744
;  PRIOR FILING DATE: 2000-11-06
;  PRIOR APPLICATION NUMBER: 60/311,966
;  PRIOR FILING DATE: 2001-08-09
;  NUMBER OF SEQ ID NOS: 47
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
;   LENGTH: 852
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant protein encoded by SEQ ID NO:24
US-10-011-588-25

  Query Match             91.3%;  Score 42;  DB 14;  Length 852;
  Best Local Similarity   87.5%;  Pred. No. 27;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||||||
Db        229 HELIHVLH 236


RESULT 11
US-10-241-596-22
; Sequence 22, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
;  APPLICANT: Microbiological Research Authority
;  APPLICANT:  The Speywood Laboratory Limited
;  TITLE OF INVENTION: Recombinant Toxin Fragments
;  FILE REFERENCE: 1581.0130003
;  CURRENT APPLICATION NUMBER: US/10/241,596
;  CURRENT FILING DATE:  2002-09-12
;  PRIOR APPLICATION NUMBER: US 09/255,829
;  PRIOR FILING DATE: 1999-02-23
;  PRIOR APPLICATION NUMBER: US 09/242,689
;  PRIOR FILING DATE: 1999-02-23
;  PRIOR APPLICATION NUMBER: PCT/GB97/02273
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: US 08/782,893
;  PRIOR FILING DATE: 1996-12-27
;  PRIOR APPLICATION NUMBER: GB 9625996.5
;  PRIOR FILING DATE: 1996-12-13
;  PRIOR APPLICATION NUMBER: GB 9617671.4
;  PRIOR FILING DATE: 1996-08-23
;  NUMBER OF SEQ ID NOS: 175
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
;   LENGTH: 858
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-241-596-22

  Query Match             91.3%;  Score 42;  DB 12;  Length 858;
  Best Local Similarity   87.5%;  Pred. No. 27;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||||||
Db        230 HELIHVLH 237


RESULT 12
US-10-241-596-175
; Sequence 175, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
;  APPLICANT: Microbiological Research Authority
;  APPLICANT:  The Speywood Laboratory Limited
;  TITLE OF INVENTION: Recombinant Toxin Fragments
;  FILE REFERENCE: 1581.0130003
;  CURRENT APPLICATION NUMBER: US/10/241,596
;  CURRENT FILING DATE:  2002-09-12
;  PRIOR APPLICATION NUMBER: US 09/255,829
;  PRIOR FILING DATE: 1999-02-23
;  PRIOR APPLICATION NUMBER: US 09/242,689
;  PRIOR FILING DATE: 1999-02-23
;  PRIOR APPLICATION NUMBER: PCT/GB97/02273
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: US 08/782,893
;  PRIOR FILING DATE: 1996-12-27
;  PRIOR APPLICATION NUMBER: GB 9625996.5
;  PRIOR FILING DATE: 1996-12-13
;  PRIOR APPLICATION NUMBER: GB 9617671.4
;  PRIOR FILING DATE: 1996-08-23
;  NUMBER OF SEQ ID NOS: 175
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 175
;   LENGTH: 860
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-241-596-175

  Query Match             91.3%;  Score 42;  DB 12;  Length 860;
  Best Local Similarity   87.5%;  Pred. No. 27;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||||||
Db        230 HELIHVLH 237


RESULT 13
US-10-241-596-94
; Sequence 94, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
;  APPLICANT: Microbiological Research Authority
```

APPLICANT: The Speywood Laboratory Limited
TITLE OF INVENTION: Recombinant Toxin Fragments
FILE REFERENCE: 1581.0130003
CURRENT APPLICATION NUMBER: US/10/241,596
CURRENT FILING DATE: 2002-09-12
PRIOR APPLICATION NUMBER: US 09/255,829
PRIOR FILING DATE: 1999-02-23
PRIOR APPLICATION NUMBER: US 09/242,689
PRIOR FILING DATE: 1999-02-23
PRIOR APPLICATION NUMBER: PCT/GB97/02273
PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 08/782,893
PRIOR FILING DATE: 1996-12-27
PRIOR APPLICATION NUMBER: GB 9625996.5
PRIOR FILING DATE: 1996-12-13
PRIOR APPLICATION NUMBER: GB 9617671.4
PRIOR FILING DATE: 1996-08-23
NUMBER OF SEQ ID NOS: 175
SOFTWARE: PatentIn version 3.1
SEQ ID NO 94
LENGTH: 862
TYPE: PRT
ORGANISM: Clostridium botulinum
US-10-241-596-94

Query Match 91.3%; Score 42; DB 12; Length 862;
Best Local Similarity 87.5%; Pred. No. 27;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HDLIHVLH 8
|:||||||
Db 232 HELIHVLH 239

RESULT 14
US-10-241-596-171
Sequence 171, Application US/10241596
Publication No. US20030166238A1
GENERAL INFORMATION:
APPLICANT: Microbiological Research Authority
APPLICANT: The Speywood Laboratory Limited
TITLE OF INVENTION: Recombinant Toxin Fragments
FILE REFERENCE: 1581.0130003
CURRENT APPLICATION NUMBER: US/10/241,596
CURRENT FILING DATE: 2002-09-12
PRIOR APPLICATION NUMBER: US 09/255,829
PRIOR FILING DATE: 1999-02-23
PRIOR APPLICATION NUMBER: US 09/242,689
PRIOR FILING DATE: 1999-02-23
PRIOR APPLICATION NUMBER: PCT/GB97/02273
PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 08/782,893
PRIOR FILING DATE: 1996-12-27
PRIOR APPLICATION NUMBER: GB 9625996.5
PRIOR FILING DATE: 1996-12-13
PRIOR APPLICATION NUMBER: GB 9617671.4
PRIOR FILING DATE: 1996-08-23
NUMBER OF SEQ ID NOS: 175
SOFTWARE: PatentIn version 3.1
SEQ ID NO 171
LENGTH: 862
TYPE: PRT
ORGANISM: Clostridium botulinum
US-10-241-596-171

Query Match 91.3%; Score 42; DB 12; Length 862;
Best Local Similarity 87.5%; Pred. No. 27;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HDLIHVLH 8
|:||||||
Db 230 HELIHVLH 237

RESULT 15
US-10-241-596-173
; Sequence 173, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
; APPLICANT: Microbiological Research Authority
; APPLICANT: The Speywood Laboratory Limited
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130003
; CURRENT APPLICATION NUMBER: US/10/241,596
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 09/255,829
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: US 09/242,689
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/GB97/02273
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 08/782,893
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB 9625996.5
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: GB 9617671.4
; PRIOR FILING DATE: 1996-08-23
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 173
; LENGTH: 862
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-241-596-173

Query Match 91.3%; Score 42; DB 12; Length 862;
Best Local Similarity 87.5%; Pred. No. 27;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HDLIHVLH 8
|:||||||
Db 230 HELIHVLH 237

Search completed: November 21, 2003, 15:58:28
Job time : 23.5 secs

GenCore version 5.1.6


M protein - protein search, using sw model

un on: November 21, 2003, 15:44:40 ; Search time 13 Seconds
(without alignments)
59.181 Million cell updates/sec

itle: US-10-064-903-2
erfect score: 46
equence: 1 HDLIHVLH 8

coring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

earched: 283308 seqs, 96168682 residues

otal number of hits satisfying chosen parameters: 283308

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| esult No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 42 | 91.3 | 1268 | 2 | S33411 | botulinum neurotox |
| 2 | 42 | 91.3 | 1291 | 1 | A48940 | bontoxilysin (EC 3 |
| 3 | 42 | 91.3 | 1291 | 2 | I40631 | non-proteolytic bo |
| 4 | 42 | 91.3 | 1297 | 2 | S39791 | neurotoxin - Clost |
| 5 | 42 | 91.3 | 1315 | 1 | BTCLTN | tentoxilysin (EC 3 |
| 6 | 38 | 82.6 | 925 | 2 | A72096 | ct234 hypothetical |
| 7 | 38 | 82.6 | 925 | 2 | E81573 | conserved hypothet |
| 8 | 38 | 82.6 | 925 | 2 | E86527 | CT234 hypothetical |
| 9 | 38 | 82.6 | 1274 | 2 | I40813 | neurotoxin type F |
| 10 | 36 | 78.3 | 232 | 2 | C85585 | unknown protein en |
| 11 | 36 | 78.3 | 232 | 2 | B90735 | hypothetical prote |
| 12 | 36 | 78.3 | 268 | 2 | S73042 | purine nucleoside |
| 13 | 36 | 78.3 | 489 | 2 | G86867 | prophage ps3 prote |
| 14 | 36 | 78.3 | 773 | 2 | C84554 | hypothetical prote |
| 15 | 36 | 78.3 | 1251 | 2 | JH0256 | botulinum neurotox |
| 16 | 36 | 78.3 | 1252 | 2 | S21178 | botulinum neurotox |
| 17 | 35 | 76.1 | 312 | 2 | C71806 | hypothetical prote |
| 18 | 35 | 76.1 | 312 | 2 | G64712 | toxR-activated gen |
| 19 | 35 | 76.1 | 416 | 2 | T45051 | hypothetical prote |
| 20 | 35 | 76.1 | 431 | 2 | T18753 | hypothetical prote |
| 21 | 35 | 76.1 | 500 | 2 | C75455 | carboxypeptidase-r |
| 22 | 35 | 76.1 | 679 | 2 | H95036 | glycosyl hydrolase |
| 23 | 35 | 76.1 | 737 | 2 | D97907 | alpha-xylosidase ( |
| 24 | 34 | 73.9 | 198 | 2 | F95194 | recombination prot |
| 25 | 34 | 73.9 | 198 | 2 | C98061 | recombination prot |
| 26 | 34 | 73.9 | 198 | 2 | AD1788 | recombination prot |
| 27 | 34 | 73.9 | 198 | 2 | AE1412 | recombination prot |
| 28 | 34 | 73.9 | 199 | 2 | JC5718 | superoxide dismuta |
| 29 | 34 | 73.9 | 421 | 2 | C84555 | hypothetical prote |
| 30 | 34 | 73.9 | 446 | 2 | B82282 | exodeoxyribonuclea |
| 31 | 34 | 73.9 | 449 | 2 | AF0820 | exodeoxyribonuclea |
| 32 | 34 | 73.9 | 456 | 1 | NCEC7 | exodeoxyribonuclea |
| 33 | 34 | 73.9 | 456 | 2 | C91050 | exonuclease VII la |
| 34 | 34 | 73.9 | 456 | 2 | H85894 | exonuclease VII, l |
| 35 | 34 | 73.9 | 523 | 2 | T04742 | hypothetical prote |
| 36 | 34 | 73.9 | 734 | 2 | D95856 | conserved hypothet |
| 37 | 34 | 73.9 | 2241 | 2 | S09811 | hypothetical prote |
| 38 | 33 | 71.7 | 29 | 2 | C60110 | repetitive protein |
| 39 | 33 | 71.7 | 186 | 2 | AF2083 | hypothetical prote |
| 40 | 33 | 71.7 | 229 | 2 | T20722 | hypothetical prote |
| 41 | 33 | 71.7 | 297 | 2 | E84237 | hypothetical prote |
| 42 | 33 | 71.7 | 352 | 2 | T38311 | protein kinase - f |
| 43 | 33 | 71.7 | 455 | 2 | H82881 | cytosol aminopepti |
| 44 | 33 | 71.7 | 459 | 2 | AH0349 | exodeoxyribonuclea |
| 45 | 33 | 71.7 | 521 | 2 | T27606 | hypothetical prote |

ALIGNMENTS

RESULT 1
S33411
botulinum neurotoxin type F - Clostridium barati
C;Species: Clostridium barati
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Jul-1999
C;Accession: S33411; S31860
R;Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, P.T.
FEMS Microbiol. Lett. 108, 175-182, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium barati type F neurotoxin
A;Reference number: S33411; MUID:93252228; PMID:8486245
A;Accession: S33411
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1268 <THO>
A;Cross-references: EMBL:X68262; NID:g49138; PIDN:CAA48329.1; PID:g49139
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 91.3%; Score 42; DB 2; Length 1268;
Best Local Similarity 87.5%; Pred. No. 6.2;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 HDLIHVLH 8
              |:||||||
Db        219 HELIHVLH 226
```

RESULT 2
A48940
bontoxilysin (EC 3.4.24.69) B precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type B (BoNT/B)
C;Species: Clostridium botulinum
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C;Accession: A48940; S48105; S21575; A42871; S07155; S08562; S07128; S08573; S08574
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Brehm, J.K.; Atkinson, T.; Minton, N.P.
Appl. Environ. Microbiol. 58, 2345-2354, 1992
A;Title: Molecular cloning of the Clostridium botulinum structural gene encoding the type
A;Reference number: A48940; MUID:92384550; PMID:1514783
A;Accession: A48940
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <WHE>
A;Cross-references: GB:M81186; NID:g144734; PIDN:AAA23211.1; PID:g144735
A;Experimental source: type B, Danish
A;Note: sequence extracted from NCBI backbone (NCBIN:112080, NCBIP:112081); this publicat
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific ide
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48105
A;Status: preliminary
A;Molecule type: DNA

;Residues: 634-994 <CAM>
.;Cross-references: EMBL:X70817; NID:g407782; PIDN:CAA50148.1; PID:g407783
.;Experimental source: proteolytic type B, strain NCTC 7273
;Szabo, E.A.; Pemberton, J.M.; Desmarchelier, P.M.
ubmitted to the EMBL Data Library, April 1992
.;Description: Partial amino acid sequence of botulinum neurotoxin type B and comparisic
.;Reference number: S21575
.;Accession: S21575
.;Molecule type: DNA
.;Residues: 36-217,'G',219-224,'S',226-246 <SZA>
.;Cross-references: EMBL:Z11934; NID:g40383; PIDN:CAA77991.1; PID:g40384
;Kurazono, H.; Mochida, S.; Binz, T.; Eisel, U.; Quanz, M.; Grebenstein, O.; Wernars, K
. Biol. Chem. 267, 14721-14729, 1992
.;Title: Minimal essential domains specifying toxicity of the light chains of tetanus to
.;Reference number: A42871; MUID:92340509; PMID:1634516
.;Accession: A42871
.;Status: nucleic acid sequence not shown
.;Molecule type: mRNA
.;Residues: 1-313,'S',315-451 <KUR>
.;Experimental source: strain Okra
.;Note: sequence extracted from NCBI backbone (NCBIP:109365)
;DasGupta, B.R.; Datta, A.
iochimie 70, 811-817, 1988
.;Title: Botulinum neurotoxin type B (strain 657): partial sequence and similarity with
.;Reference number: S07155; MUID:89000987; PMID:3139097
.;Accession: S07155
.;Molecule type: protein
.;Residues: 2-29,'M',31-45 <DAS>
.;Accession: S08562
.;Molecule type: protein
.;Residues: 442-463,'R',465-467 <DA2>
;Schmidt, J.J.; Sathyamoorthy, V.; DasGupta, B.R.
rch. Biochem. Biophys. 238, 544-548, 1985
.;Title: Partial amino acid sequences of botulinum neurotoxins types B and E.
.;Reference number: S07128; MUID:85197963; PMID:3888113
.;Accession: S07128
.;Status: preliminary
.;Molecule type: protein
.;Residues: 2-16 <SCH1>
.;Accession: S08573
.;Status: preliminary
.;Molecule type: protein
.;Residues: 2-17 <SCH2>
.;Accession: S08574
.;Status: preliminary
.;Molecule type: protein
.;Residues: 442-459 <SCH3>
;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R
lature 359, 832-835, 1992
.;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyt
.;Reference number: S27125; MUID:93063293; PMID:1331807
.;Contents: annotation
;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synap
;Genetics:
.;Gene: bont/b
;Function:
.;Description: catalyzes hydrolysis of a Gln-Phe peptide bond in synaptobrevin 2
;Superfamily: tetanus toxin
;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
;2-441/Product: bontoxilysin B light chain #status experimental <LGHT>
;442-1291/Product: bontoxilysin B heavy chain #status experimental <HVY>
;230,234/Binding site: zinc (His) #status predicted
;231/Active site: Glu #status predicted

  Query Match 91.3%; Score 42; DB 1; Length 1291;
  Best Local Similarity 87.5%; Pred. No. 6.3;
  Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

)y 1 HDLIHVLH 8
|:||||||
)b 230 HELIHVLH 237

RESULT 3
I40631
non-proteolytic botulinum neurotoxin type B precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40631; S48103; S48104; S36015
R;Hutson, R.A.; Collins, M.D.; East, A.K.; Thompson, D.E.
Curr. Microbiol. 28, 101-110, 1994
A;Title: Nucleotide sequence of the gene coding for non-proteolytic Clostridium botulinum
A;Reference number: I40631; MUID:94122659; PMID:7764370
A;Accession: I40631
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1291 <RES>
A;Cross-references: EMBL:X71343; NID:g296148; PIDN:CAA50482.1; PID:g296149
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific ide
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48103
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 634-761,'E',763-841,'M',843,'T',845,'N',847-994 <CAM1>
A;Cross-references: EMBL:X70814; NID:g407778; PIDN:CAA50145.1; PID:g407779
A;Experimental source: non-proteolytic strain 2129B (Scott)
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
A;Accession: S48104
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-843,'T',845,'N',847-994 <CAM2>
A;Cross-references: EMBL:X70819; NID:g407780; PIDN:CAA50150.1; PID:g407781
A;Experimental source: non-proteolytic strain Eklund 2B (Colworth 229)
C;Comment: Botulinum neurotoxin type B in these strains may posses a capable catalytic si
C;Genetics:
A;Gene: bont/b
C;Superfamily: tetanus toxin
C;Keywords: metalloprotein; neurotoxin; transmembrane protein; zinc
F;2-441/Product: botulinum neurotoxin type B light chain #status predicted <LGHT>
F;442-1291/Product: botulinum neurotoxin type B heavy chain #status predicted <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

  Query Match 91.3%; Score 42; DB 2; Length 1291;
  Best Local Similarity 87.5%; Pred. No. 6.3;
  Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HDLIHVLH 8
|:||||||
Db 230 HELIHVLH 237

RESULT 4
S39791
neurotoxin - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 16-Jul-1999
C;Accession: S39791
R;Campbell, K.; Collins, M.D.; East, A.K.
Biochim. Biophys. Acta 1216, 487-491, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium botulinum (Clostridium a
A;Reference number: S39791; MUID:94092745; PMID:8268233
A;Accession: S39791
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1297 <CAM>
A;Cross-references: EMBL:X74162; NID:g441275; PIDN:CAA52275.1; PID:g441276
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match 91.3%; Score 42; DB 2; Length 1297;
  Best Local Similarity 87.5%; Pred. No. 6.3;
  Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 HDLIHVLH 8
              |:||||||
Db        230 HELIHVLH 237
```

ESULT 5
TCLTN
entoxilysin (EC 3.4.24.68) precursor - Clostridium tetani
;Alternate names: tetanus neurotoxin
;Species: Clostridium tetani
;Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 03-Jun-2002
;Accession: A25689; A25757; A25194; B25194; A60759; S69348; S09364
;Eisel, U.; Jarausch, W.; Goretzki, K.; Henschen, A.; Engels, J.; Weller, U.; Hudel, M.
MBO J. 5, 2495-2502, 1986
;Title: Tetanus toxin: primary structure, expression in E. coli, and homology with botu
;Reference number: A25689; MUID:87053814; PMID:3536478
;Accession: A25689
;Molecule type: DNA
;Residues: 1-1315 <EIS>
;Cross-references: GB:X04436; NID:g40769; PIDN:CAA28033.1; PID:g40770
;Fairweather, N.F.; Lyness, V.A.
ucleic Acids Res. 14, 7809-7812, 1986
;Title: The complete nucleotide sequence of tetanus toxin.
;Reference number: A25757; MUID:87040747; PMID:3774547
;Accession: A25757
;Molecule type: DNA
;Residues: 1-1315 <FAI>
;Cross-references: GB:X06214; NID:g40773; PIDN:CAA29564.1; PID:g40774
;Experimental source: strain CN3911
;Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.; Thomson, R.O.
. Bacteriol. 165, 21-27, 1986
;Title: Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in E
;Reference number: A25194; MUID:86085672; PMID:3510187
;Accession: A25194
;Molecule type: DNA
;Residues: 743-1315 <FA2>
;Cross-references: GB:M12739; NID:g144920; PIDN:AAA23282.1; PID:g144921
;Accession: B25194
;Molecule type: protein
;Residues: 865-894 <FA3>
;Matsuda, M.; Lei, D.L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.
nfect. Immun. 57, 3588-3593, 1989
;Title: Isolation, purification, and characterization of fragment B, the NH-2-terminal
;Reference number: A60759; MUID:90035436; PMID:2478476
;Accession: A60759
;Molecule type: protein
;Residues: 461-475 <MAT>
;Demotz, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.; Widmann, C.; Corradin, G.
. Immunol. 142, 394-402, 1989
;Title: Delineation of several DR-restricted tetanus toxin T cell epitopes.
;Reference number: JS0098; MUID:89093918; PMID:2463305
;Contents: annotation; epitope region
;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R
ature 359, 832-835, 1992
;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyt
;Reference number: S27125; MUID:93063293; PMID:1331807
;Contents: annotation
;de Filippis, V.; Vangelista, L.; Schiavo, G.; Tonello, F.; Montecucco, C.
ur. J. Biochem. 229, 61-69, 1995
;Title: Structural studies on the zinc-endopeptidase light chain of tetanus neurotoxin.
;Reference number: S69348; MUID:95262688; PMID:7744050
;Accession: S69348
;Molecule type: protein
;Residues: 2-31 <DEF>
;Comment: The source of this protein was an extrachromosomal plasmid.
;Comment: The precursor is cleaved by endogenous proteinase activity to form light (fra
ual chains are not toxic when separated). The amino end of the heavy chain (fragment B)
;Comment: Fragment B forms ion channels in a lipid bilayer. Fragment C binds to ganglic
;Comment: This potent neurotoxin binds to peripheral neuronal synapses, is internalized
presynaptic neurons. It inhibits neurotransmitter release by proteolytic cleavage of sy
;Function:
;Description: blocks neuroexocytosis via hydrolysis of a Gln-Phe peptide bond in synapt
;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-457/Product: tentoxylysin light chain (fragment A) #status predicted <TTL>
F;461-1315/Product: tentoxylysin heavy chain (fragment B.C) #status experimental <TTH>
F;461-864/Domain: channel forming (fragment B) #status predicted <TXB>
F;865-1315/Domain: ganglioside binding (fragment C) #status predicted <TXC>
F;233,237/Binding site: zinc (His) #status predicted
F;234/Active site: Glu #status predicted

  Query Match  91.3%;  Score 42;  DB 1;  Length 1315;
  Best Local Similarity  87.5%;  Pred. No. 6.4;
  Matches  7;  Conservative  1;  Mismatches  0;  Indels  0;  Gaps  0;

```
Qy          1 HDLIHVLH 8
              |:||||||
Db        233 HELIHVLH 240
```

RESULT 6
A72096
ct234 hypothetical protein - Chlamydophila pneumoniae (strain CWL029)
C;Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Jun-2000
C;Accession: A72096
R;Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A;Title: Comparative genomes of Clamydia pneumoniae and C. trachomatis.
A;Reference number: A72000; MUID:99206606; PMID:10192388
A;Accession: A72096
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-925 <ARN>
A;Cross-references: GB:AE001614; GB:AE001363; NID:g4376562; PIDN:AAD18442.1; PID:g4376568
A;Experimental source: strain CWL029
C;Genetics:
A;Gene: CPn0293
C;Superfamily: Chlamydia trachomatis hypothetical protein CT234

  Query Match  82.6%;  Score 38;  DB 2;  Length 925;
  Best Local Similarity  62.5%;  Pred. No. 25;
  Matches  5;  Conservative  2;  Mismatches  1;  Indels  0;  Gaps  0;

```
Qy          1 HDLIHVLH 8
              |||:|: |
Db        538 HDLLHITH 545
```

RESULT 7
E81573
conserved hypothetical protein CP0465 [imported] - Chlamydophila pneumoniae (strain AR39)
C;Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 20-Jun-2000
C;Accession: E81573
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, I
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A;Reference number: A81500; MUID:20150255; PMID:10684935
A;Accession: E81573
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-925 <REA>
A;Cross-references: GB:AE002208; GB:AE002161; NID:g7189387; PIDN:AAF38302.1; PID:g7189388
A;Experimental source: strain AR39, HL cells
C;Genetics:
A;Gene: CP0465
C;Superfamily: Chlamydia trachomatis hypothetical protein CT234

  Query Match  82.6%;  Score 38;  DB 2;  Length 925;
  Best Local Similarity  62.5%;  Pred. No. 25;
  Matches  5;  Conservative  2;  Mismatches  1;  Indels  0;  Gaps  0;

```
Qy          1 HDLIHVLH 8
              |||:|: |
```

```
b        538 HDLLHITH 545


ESULT 8
86527
T234 hypothetical protein [imported] - Chlamydophila pneumoniae (strain J138)
;Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
;Accession: E86527
;Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Is
ucleic Acids Res. 28, 2311-2314, 2000
.;Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
.;Reference number: A86491; MUID:20330349; PMID:10871362
.;Accession: E86527
.;Status: preliminary
.;Molecule type: DNA
.;Residues: 1-925 <STO>
.;Cross-references: GB:BA000008; NID:g8978667; PIDN:BAA98503.1; GSPDB:GN00142
.;Experimental source: strain J138
';Genetics:
.;Gene: CPj0293
';Superfamily: Chlamydia trachomatis hypothetical protein CT234

  Query Match             82.6%;  Score 38;  DB 2;  Length 925;
  Best Local Similarity   62.5%;  Pred. No. 25;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

)y          1 HDLIHVLH 8
              |||:|: |
)b        538 HDLLHITH 545


:ESULT 9
 40813
ieurotoxin type F - Clostridium botulinum
';Species: Clostridium botulinum
';Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 16-Jul-1999
';Accession: I40813; S48108
';East, A.K.; Richardson, P.T.; Allaway, D.; Collins, M.D.; Roberts, T.A.; Thompson, D.E
'EMS Microbiol. Lett. 96, 225-230, 1992
\;Title: Sequence of the gene encoding type F neurotoxin of Clostridium botulinum.
\;Reference number: I40644
\;Accession: I40813
\;Status: preliminary; translated from GB/EMBL/DDBJ
\;Molecule type: DNA
\;Residues: 1-1274 <RES>
\;Cross-references: GB:M92906; NID:g144866; PIDN:AAA23263.1; PID:g144867
';Campbell, K.D.; Collins, M.D.; East, A.K.
'. Clin. Microbiol. 31, 2255-2262, 1993
\;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
\;Reference number: S48103; MUID:94013372; PMID:8408542
\;Accession: S48108
\;Status: preliminary; translation not shown
\;Molecule type: DNA
\;Residues: 634-1002 <CAM>
\;Cross-references: EMBL:X70816; NID:g407788; PIDN:CAA50147.1; PID:g407789
';Superfamily: tetanus toxin
';Keywords: neurotoxin

  Query Match             82.6%;  Score 38;  DB 2;  Length 1274;
  Best Local Similarity   75.0%;  Pred. No. 36;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

)y          1 HDLIHVLH 8
              |:||| ||
)b        227 HELIHALH 234


RESULT 10
C85585
inknown protein encoded by prophage CP-933K [imported] - Escherichia coli (strain O157:H
';Species: Escherichia coli
';Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: C85585
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca.
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: C85585
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-232 <STO>
A;Cross-references: GB:AE005174; NID:g12513758; PIDN:AAG55143.1; GSPDB:GN00145; UWGP:Z09
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z0990

  Query Match             78.3%;  Score 36;  DB 2;  Length 232;
  Best Local Similarity   62.5%;  Pred. No. 14;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:|:|| |
Db        142 HELLHVFH 149


RESULT 11
B90735
hypothetical protein ECs0850 [imported] - Escherichia coli (strain O157:H7, substrain RIM
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: B90735
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: B90735
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-232 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB34273.1; PID:g13360309; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs0850

  Query Match             78.3%;  Score 36;  DB 2;  Length 232;
  Best Local Similarity   62.5%;  Pred. No. 14;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:|:|| |
Db        142 HELLHVFH 149


RESULT 12
S73042
purine nucleoside phosphorylase pnpH - Mycobacterium leprae
N;Alternate names: L308_F2_56 protein
C;Species: Mycobacterium leprae
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 23-Mar-2001
C;Accession: S73042
R;Smith, D.R.; Robison, K.
submitted to the EMBL Data Library, November 1993
A;Description: Mycobacterium leprae cosmid L308.
A;Reference number: S72590
A;Accession: S73042
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-268 <SMI>
A;Cross-references: EMBL:U00022; NID:g467164; PIDN:AAA17341.1; PID:g467183
C;Genetics:
A;Gene: pnpH
A;Start codon: GTG
C;Superfamily: purine-nucleoside phosphorylase
```

```
 Query Match             78.3%;  Score 36;  DB 2;  Length 268;
 Best Local Similarity   75.0%;  Pred. No. 16;
 Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              ||| ||:|
Db         95 HDLRHVVH 102


RESULT 13
S86867
prophage ps3 protein 11 [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C;Accession: G86867
A;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrli
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: G86867
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-489 <STO>
A;Cross-references: GB:AE005176; PID:g12724983; PIDN:AAK06041.1; GSPDB:GN00146
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: ps311

 Query Match             78.3%;  Score 36;  DB 2;  Length 489;
 Best Local Similarity   71.4%;  Pred. No. 31;
 Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 DLIHVLH 8
              ||:|:||
Db         49 DLVHILH 55


RESULT 14
S84554
hypothetical protein At2g17610 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: C84554
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
1.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
Fuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: C84554
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-773 <STO>
A;Cross-references: GB:AE002093; NID:g4926870; PIDN:AAD32950.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g17610
A;Map position: 2

 Query Match             78.3%;  Score 36;  DB 2;  Length 773;
 Best Local Similarity   75.0%;  Pred. No. 50;
 Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||| ||
Db         10 HELIHSLH 17


RESULT 15
JH0256
botulinum neurotoxin type E precursor - Clostridium butyricum
C;Species: Clostridium butyricum
C;Date: 30-Jun-1992 #sequence_revision 15-May-1998 #text_change 16-Jul-1999
C;Accession: JH0256; S16145
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428; PMID:1543481
A;Accession: JH0256
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-27,'E',29-1251 <POU>
A;Cross-references: EMBL:X62088; NID:g40379
A;Experimental source: strains ATCC 43181 and ATCC 43755
R;Fujii, N.; Kimura, K.; Yashiki, T.; Indoh, T.; Murakami, T.; Tsuzuki, K.; Yokosawa, N.;
J. Gen. Microbiol. 137, 519-525, 1991
A;Title: Cloning of a DNA fragment encoding the 5'-terminus of the botulinum type E toxi
A;Reference number: S16145; MUID:91237316; PMID:2033376
A;Accession: S16145
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-229,'M',231-252 <FUJ>
A;Cross-references: EMBL:X53180; NID:g40407; PIDN:CAA37321.1; PID:g40408
A;Experimental source: strain BL6340
C;Comment: The clostridial neurotoxins are toxins that inhibit neurotransmitter release
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the ligh
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;2-422/Product: botulinum neurotoxin type E light chain #status predicted <LIG>
F;423-1251/Product: botulinum neurotoxin type E heavy chain #status predicted <HEA>
F;412-426/Disulfide bonds: #status predicted

  Query Match             78.3%;  Score 36;  DB 2;  Length 1251;
  Best Local Similarity   75.0%;  Pred. No. 85;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||| ||
Db        212 HELIHSLH 219


Search completed: November 21, 2003, 15:50:23
Job time : 13 secs
```

GenCore version 5.1.6


)M protein - protein search, using sw model

tun on: November 21, 2003, 15:30:05 ; Search time 10 Seconds
(without alignments)
37.621 Million cell updates/sec

:itle: US-10-064-903-2
?erfect score: 46
Sequence: 1 HDLIHVLH 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 42 | 91.3 | 1290 | 1 | BXB_CLOBO | P10844 clostridium |
| 2 | 42 | 91.3 | 1296 | 1 | BXG_CLOBO | Q60393 clostridium |
| 3 | 42 | 91.3 | 1314 | 1 | TETX_CLOTE | P04958 clostridium |
| 4 | 38 | 82.6 | 1274 | 1 | BXF_CLOBO | P30996 clostridium |
| 5 | 36 | 78.3 | 268 | 1 | PUNA_MYCLE | P46862 mycobacteri |
| 6 | 36 | 78.3 | 1250 | 1 | BXE_CLOBO | Q00496 clostridium |
| 7 | 36 | 78.3 | 1250 | 1 | BXE_CLOBU | P30995 clostridium |
| 8 | 34 | 73.9 | 198 | 1 | RECR_LISIN | Q927d9 listeria in |
| 9 | 34 | 73.9 | 198 | 1 | RECR_LISMO | Q8y3x7 listeria mo |
| 10 | 34 | 73.9 | 198 | 1 | RECR_STRPN | Q9zhc4 streptococc |
| 11 | 34 | 73.9 | 199 | 1 | SODC_HAEDU | Q59452 haemophilus |
| 12 | 34 | 73.9 | 446 | 1 | EX7L_VIBCH | Q9ktw4 vibrio chol |
| 13 | 34 | 73.9 | 449 | 1 | EX7L_SALTI | Q8z4q1 salmonella |
| 14 | 34 | 73.9 | 449 | 1 | EX7L_SALTY | Q8zn58 salmonella |
| 15 | 34 | 73.9 | 456 | 1 | EX7L_ECO57 | Q8xab0 escherichia |
| 16 | 34 | 73.9 | 456 | 1 | EX7L_ECOLI | P04994 escherichia |
| 17 | 34 | 73.9 | 458 | 1 | EX7L_ECOL6 | Q8ff64 escherichia |
| 18 | 34 | 73.9 | 2241 | 1 | TEGU_HCMVA | P16785 human cytom |
| 19 | 33 | 71.7 | 239 | 1 | CENB_SHEEP | P49451 ovis aries |
| 20 | 33 | 71.7 | 352 | 1 | PRK1_SCHPO | O13958 schizosacch |
| 21 | 33 | 71.7 | 459 | 1 | EX7L_YERPE | Q8zcu2 yersinia pe |
| 22 | 33 | 71.7 | 541 | 1 | UME5_YEAST | P39073 saccharomyc |
| 23 | 33 | 71.7 | 599 | 1 | CENB_HUMAN | P07199 homo sapien |
| 24 | 33 | 71.7 | 599 | 1 | CENB_MOUSE | P27790 mus musculu |
| 25 | 33 | 71.7 | 606 | 1 | CENB_CRIGR | P48988 cricetulus |
| 26 | 33 | 71.7 | 697 | 1 | AD26_MOUSE | Q9r158 mus musculu |
| 27 | 33 | 71.7 | 740 | 1 | FAS_PNECA | P29251 p folic aci |
| 28 | 33 | 71.7 | 942 | 1 | AMPN_MANSE | P91685 manduca sex |
| 29 | 32 | 69.6 | 91 | 1 | YVBG_VACCC | P20547 vaccinia vi |
| 30 | 32 | 69.6 | 239 | 1 | MT04_ECOLI | P32049 escherichia |
| 31 | 32 | 69.6 | 477 | 1 | GLGA_CLOAB | Q97gx6 clostridium |
| 32 | 32 | 69.6 | 528 | 1 | UGA3_YEAST | P26370 saccharomyc |
| 33 | 32 | 69.6 | 612 | 1 | OCTC_RAT | P11466 rattus norv |
| 34 | 32 | 69.6 | 832 | 1 | DPO1_THEAQ | P19821 thermus aqu |
| 35 | 32 | 69.6 | 918 | 1 | BARA_ECOLI | P26607 escherichia |
| 36 | 32 | 69.6 | 918 | 1 | BARA_SHIFL | P59342 shigella fl |
| 37 | 32 | 69.6 | 970 | 1 | PSU1_YEAST | P53550 saccharomyc |
| 38 | 32 | 69.6 | 2524 | 1 | NOTC_XENLA | P21783 xenopus lae |
| 39 | 31 | 67.4 | 110 | 1 | YZ15_AQUAE | O66407 aquifex aeo |
| 40 | 31 | 67.4 | 268 | 1 | COQ4_DROME | Q9vvg6 drosophila |
| 41 | 31 | 67.4 | 286 | 1 | HTPX_RALSO | Q8y3a6 ralstonia s |
| 42 | 31 | 67.4 | 304 | 1 | HTX1_STRCO | Q9rkn3 streptomyce |
| 43 | 31 | 67.4 | 336 | 1 | HTPX_RHILO | Q98et0 rhizobium l |
| 44 | 31 | 67.4 | 426 | 1 | SYW_THEAC | Q9hiw5 thermoplasm |
| 45 | 31 | 67.4 | 444 | 1 | TRME_CHLMU | Q9plm9 chlamydia m |

ALIGNMENTS

RESULT 1
BXB_CLOBO
ID BXB_CLOBO STANDARD; PRT; 1290 AA.
AC P10844; P10843;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Botulinum neurotoxin type B precursor (EC 3.4.24.69) (BoNT/B)
DE (Bontoxilysin B).
GN BOTB.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92384550; PubMed=1514783;
RA Whelan S.M., Elmore M.J., Bodsworth N.J., Brehm J.K., Atkinson T.,
RA Minton N.P.;
RT "Molecular cloning of the Clostridium botulinum structural gene
RT encoding the type B neurotoxin and determination of its entire
RT nucleotide sequence.";
RL Appl. Environ. Microbiol. 58:2345-2354(1992).
RN [2]
RP SEQUENCE OF 35-245 FROM N.A.
RC STRAIN=NCTC 7273;
RA Szabo E.A., Pemberton J.M., Desmarchelier P.M.;
RL Submitted (APR-1992) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE OF 633-993 FROM N.A.
RC STRAIN=NCTC 7273;
RX MEDLINE=94013372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulinal neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.";
RL J. Clin. Microbiol. 31:2255-2262(1993).
RN [4]
RP SEQUENCE OF 1-44 AND 441-466.
RC STRAIN=657;
RX MEDLINE=89000987; PubMed=3139097;
RA Dasgupta B.R., Datta A.;
RT "Botulinum neurotoxin type B (strain 657): partial sequence and
RT similarity with tetanus toxin.";
RL Biochimie 70:811-817(1988).
RN [5]
RP SEQUENCE OF 1-16 AND 441-458.
RC STRAIN=OKRA;
RX MEDLINE=85197963; PubMed=3888113;
RA Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT "Partial amino acid sequences of botulinum neurotoxins types B and
RT E.";
RL Arch. Biochem. Biophys. 238:544-548(1985).
RN [6]
RP IDENTIFICATION AS ZINC-PROTEASE.
RX MEDLINE=93054694; PubMed=1429690;
RA Schiavo G., Rossetto O., Santucci A., Dasgupta B.R., Montecucco C.;

```
RT   "Botulinum neurotoxins are zinc proteins.";
RL   J. Biol. Chem. 267:23479-23483(1992).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=93063293; PubMed=1331807;
RA   Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA   Dasgupta B.R., Montecucco C.;
RT   "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT   by proteolytic cleavage of synaptobrevin.";
RL   Nature 359:832-835(1992).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES THE 76-GLN-|-PHE-77 BOND OF
CC       SYNAPTOBREVIN-2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC       HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC       WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC       FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC       BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81186; AAA23211.1; -.
DR   EMBL; Z11934; CAA77991.1; -.
DR   EMBL; X70817; CAA50148.1; -.
DR   PIR; A48940; A48940.
DR   PDB; 1EPW; 01-NOV-00.
DR   PDB; 1F31; 01-NOV-00.
DR   PDB; 1F82; 16-AUG-00.
DR   PDB; 1F83; 16-AUG-00.
DR   PDB; 1FQH; 06-DEC-00.
DR   PDB; 1G9A; 13-NOV-02.
DR   PDB; 1G9B; 13-NOV-02.
DR   PDB; 1G9C; 13-NOV-02.
DR   PDB; 1G9D; 13-NOV-02.
DR   PDB; 1I1E; 21-NOV-01.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR006025; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    440       BOTULINUM NEUROTOXIN B, LIGHT-CHAIN.
FT   CHAIN       441   1290       BOTULINUM NEUROTOXIN B, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    445       INTERCHAIN (PROBABLE).
FT   CONFLICT     29     29       T -> M (IN REF. 4).
FT   CONFLICT    217    217       R -> G (IN REF. 2).
FT   CONFLICT    224    224       A -> S (IN REF. 2).
FT   CONFLICT    463    463       S -> R (IN REF. 4).
SQ   SEQUENCE   1290 AA;  150670 MW;  D21746E2C024DF43 CRC64;
```

```
  Query Match             91.3%;  Score 42;  DB 1;  Length 1290;
  Best Local Similarity   87.5%;  Pred. No. 2.8;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||||||
Db        229 HELIHVLH 236


RESULT 2
BXG_CLOBO
ID   BXG_CLOBO      STANDARD;      PRT;  1296 AA.
AC   Q60393;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type G precursor (EC 3.4.24.69) (BoNT/G)
DE   (Bontoxilysin G).
GN   BOTG.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=113 / 30;
RX   MEDLINE=94092745; PubMed=8268233;
RA   Campbell K., Collins M.D., East A.K.;
RT   "Nucleotide sequence of the gene coding for Clostridium botulinum
RT   (Clostridium argentinense) type G neurotoxin: genealogical comparison
RT   with other clostridial neurotoxins.";
RL   Biochim. Biophys. Acta 1216:487-491(1993).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC       BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X74162; CAA52275.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR006025; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    441       BOTULINUM NEUROTOXIN G, LIGHT-CHAIN.
FT   CHAIN       442   1296       BOTULINUM NEUROTOXIN G, HEAVY-CHAIN.
```

```
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    435    449       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1296 AA;  149013 MW;  DC8E47E15F665C31 CRC64;

  Query Match             91.3%;  Score 42;  DB 1;  Length 1296;
  Best Local Similarity   87.5%;  Pred. No. 2.8;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||||||
Db        229 HELIHVLH 236


RESULT 3
TETX_CLOTE
ID   TETX_CLOTE     STANDARD;      PRT;  1314 AA.
AC   P04958;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Tetanus toxin precursor (EC 3.4.24.68) (Tentoxylysin) [Contains:
DE   Tetanus toxin light chain (Tetanus toxin chain L); Tetanus toxin heavy
DE   chain (Tetanus toxin chain H)].
GN   TETX OR CTP60.
OS   Clostridium tetani.
OG   Plasmid pE88, and Plasmid 75 Kbp.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1513;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   PLASMID=75 Kbp;
RX   MEDLINE=87053814; PubMed=3536478;
RA   Eisel U., Jarausch W., Goretzki K., Henschen A., Engels J.,
RA   Weller U., Hudel M., Habermann E., Niemann H.;
RT   "Tetanus toxin: primary structure, expression in E. coli, and
RT   homology with botulinum toxins.";
RL   EMBO J. 5:2495-2502(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CN3911; PLASMID=75 Kbp;
RX   MEDLINE=87040747; PubMed=3774547;
RA   Fairweather N.F., Lyness V.A.;
RT   "The complete nucleotide sequence of tetanus toxin.";
RL   Nucleic Acids Res. 14:7809-7812(1986).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Massachusetts / E88; PLASMID=pE88;
RX   MEDLINE=22457253; PubMed=12552129;
RA   Brueggemann H., Baeumer S., Fricke W.F., Wiezer A., Liesegang H.,
RA   Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA   Gottschalk G.;
RT   "The genome sequence of Clostridium tetani, the causative agent of
RT   tetanus disease.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
RN   [4]
RP   SEQUENCE OF 742-1314 FROM N.A.
RC   PLASMID=75 Kbp;
RX   MEDLINE=86085672; PubMed=3510187;
RA   Fairweather N.F., Lyness V.A., Pickard D.J., Allen G., Thomson R.O.;
RT   "Cloning, nucleotide sequencing, and expression of tetanus toxin
RT   fragment C in Escherichia coli.";
RL   J. Bacteriol. 165:21-27(1986).
RN   [5]
RP   PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX   MEDLINE=90201034; PubMed=2108021;
RA   Krieglstein K., Henschen A., Weller U., Habermann E.;
RT   "Arrangement of disulfide bridges and positions of sulfhydryl groups
RT   in tetanus toxin.";
RL   Eur. J. Biochem. 188:39-45(1990).
RN   [6]
RP   PARTIAL SEQUENCE.
RX   MEDLINE=92037649; PubMed=1935979;
RA   Krieglstein K.G., Henschen A.H., Weller U., Habermann E.;
RT   "Limited proteolysis of tetanus toxin. Relation to activity and
RT   identification of cleavage sites.";
RL   Eur. J. Biochem. 202:41-51(1991).
RN   [7]
RP   IDENTIFICATION AS ZINC-PROTEASE.
RX   MEDLINE=93010948; PubMed=1396558;
RA   Schiavo G., Poulain B., Rossetto O., Benfenati F., Tauc L.,
RA   Montecucco C.;
RT   "Tetanus toxin is a zinc protein and its inhibition of
RT   neurotransmitter release and protease activity depend on zinc.";
RL   EMBO J. 11:3577-3583(1992).
RN   [8]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=93063293; PubMed=1331807;
RA   Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA   DasGupta B.R., Montecucco C.;
RT   "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT   by proteolytic cleavage of synaptobrevin.";
RL   Nature 359:832-835(1992).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
RX   MEDLINE=97475217; PubMed=9334741;
RA   Umland T.C., Wingert L.M., Swaminathan S., Furey W.F., Schmidt J.J.,
RA   Sax M.;
RT   "Structure of the receptor binding fragment HC of tetanus
RT   neurotoxin.";
RL   Nat. Struct. Biol. 4:788-792(1997).
CC   -!- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77
CC       BOND OF SYNAPTOBREVIN-2.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of 76-Gln-|-Phe-77 bond in
CC       synaptobrevin 2.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
CC       YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
CC       AND ARE NON-TOXIC AFTER SEPARATION.
CC   -!- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO
CC       GANGLIOSIDE RECEPTORS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X04436; CAA28033.1; -.
DR   EMBL; X06214; CAA29564.1; -.
DR   EMBL; AF528097; AAO37454.1; -.
DR   EMBL; M12739; AAA23282.1; -.
DR   PIR; A25689; BTCLTN.
DR   PDB; 1AF9; 29-APR-98.
DR   PDB; 1A8D; 14-OCT-98.
DR   PDB; 1D0H; 27-MAR-00.
DR   PDB; 1DFQ; 24-MAR-00.
DR   PDB; 1DIW; 24-MAR-00.
DR   PDB; 1DLL; 24-MAR-00.
DR   PDB; 1FV3; 05-SEP-01.
DR   MEROPS; M27.001; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR006025; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
```

```
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
CW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; Plasmid;
CW   3D-structure; Complete proteome.
FT   INIT_MET      0      0
FT   CHAIN         1    456       TETANUS TOXIN LIGHT CHAIN.
FT   CHAIN       457   1314       TETANUS TOXIN HEAVY CHAIN.
FT   METAL       232    232       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    233    233       BY SIMILARITY.
FT   METAL       236    236       ZINC (CATALYTIC) (BY SIMILARITY).
FT   TRANSMEM    226    246       POTENTIAL.
FT   TRANSMEM    669    689       POTENTIAL.
FT   DISULFID    438    466       INTERCHAIN.
FT   DISULFID   1076   1092
FT   HELIX       876    882
FT   TURN        883    883
FT   STRAND      884    891
FT   TURN        892    893
FT   STRAND      894    897
FT   STRAND      904    907
FT   TURN        909    910
FT   STRAND      912    915
FT   STRAND      920    925
FT   TURN        928    929
FT   STRAND      932    935
FT   HELIX       938    940
FT   TURN        941    946
FT   STRAND      949    956
FT   HELIX       962    968
FT   TURN        969    970
FT   STRAND      972    977
FT   STRAND      980    981
FT   HELIX       983    985
FT   STRAND      987    995
FT   TURN        996    997
FT   STRAND      998   1004
FT   TURN       1006   1007
FT   STRAND     1010   1016
FT   STRAND     1020   1020
FT   TURN       1021   1022
FT   STRAND     1031   1037
FT   TURN       1039   1040
FT   STRAND     1042   1047
FT   TURN       1048   1049
FT   STRAND     1050   1056
FT   TURN       1058   1059
FT   STRAND     1068   1074
FT   TURN       1079   1080
FT   STRAND     1082   1091
FT   HELIX      1097   1105
FT   TURN       1106   1107
FT   STRAND     1112   1112
FT   STRAND     1114   1114
FT   TURN       1116   1117
FT   STRAND     1120   1120
FT   STRAND     1122   1122
FT   TURN       1123   1124
FT   STRAND     1127   1131
FT   HELIX      1132   1134
FT   TURN       1135   1136
FT   STRAND     1137   1141
FT   TURN       1144   1145
FT   STRAND     1148   1152
FT   STRAND     1155   1158
FT   TURN       1159   1162
FT   STRAND     1163   1166
FT   STRAND     1173   1178
FT   TURN       1184   1185
FT   STRAND     1188   1188
FT   STRAND     1190   1190
FT   TURN       1191   1192
FT   STRAND     1193   1201

  Query Match          91.3%;  Score 42;  DB 1;  Length 1314;
  Best Local Similarity   87.5%;  Pred. No. 2.8;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||||||
Db        232 HELIHVLH 239


RESULT 4
BXF_CLOBO
ID   BXF_CLOBO      STANDARD;      PRT;  1274 AA.
AC   P30996;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type F precursor (EC 3.4.24.69) (BoNT/F)
DE   (Bontoxilysin F).
GN   BOTF.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 23387;
RX   MEDLINE=93012902; PubMed=1398040;
RA   East A.K., Richardson P.T., Allaway D., Collins M.D.,
RA   Roberts T.A., Thompson D.E.;
RT   "Sequence of the gene encoding type F neurotoxin of Clostridium
RT   botulinum.";
RL   FEMS Microbiol. Lett. 75:225-230(1992).
RN   [2]
RP   SEQUENCE OF 1-64 FROM N.A.
RC   STRAIN=Hobbs FT10;
RX   MEDLINE=94297488; PubMed=7764998;
RA   East A.K., Collins M.D.;
RT   "Conserved structure of genes encoding components of botulinum
RT   neurotoxin complex M and the sequence of the gene coding for the
RT   nontoxic component in nonproteolytic Clostridium botulinum type F.";
RL   Curr. Microbiol. 29:69-77(1994).
RN   [3]
RP   SEQUENCE OF 634-1002 FROM N.A.
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94230352; PubMed=8175689;
RA   Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA   Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT   "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT   F botulinal neurotoxins and tetanus toxin.";
RL   J. Biol. Chem. 269:12764-12772(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 58-GLN-|-LYS-59
CC       BOND OF SYNAPTOBREVINS-1 AND -2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC       HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC       WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC       FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC       BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
```

```
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M92906; AAA23263.1; -.
DR   EMBL; S73676; AAC60475.1; -.
DR   EMBL; X70820; CAA50151.1; -.
DR   EMBL; X70816; CAA50147.1; -.
DR   PIR; I40813; I40813.
DR   PIR; S48109; S48109.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR006025; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   CHAIN         1    436       BOTULINUM NEUROTOXIN F, LIGHT-CHAIN.
FT   CHAIN       437   1274       BOTULINUM NEUROTOXIN F, HEAVY-CHAIN.
FT   METAL       227    227       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    228    228       BY SIMILARITY.
FT   METAL       231    231       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    445       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1274 AA;  146709 MW;  5B99756A7438B921 CRC64;

  Query Match             82.6%;  Score 38;  DB 1;  Length 1274;
  Best Local Similarity   75.0%;  Pred. No. 16;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||| ||
Db        227 HELIHALH 234


RESULT 5
PUNA_MYCLE
ID   PUNA_MYCLE     STANDARD;      PRT;   268 AA.
AC   P46862;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Purine nucleoside phosphorylase (EC 2.4.2.1) (Inosine phosphorylase)
DE   (PNP).
GN   PUNA OR DEOD OR ML0707 OR L308_F2_56.
OS   Mycobacterium leprae.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1769;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Smith D.R., Robison K.;
RL   Submitted (MAR-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TN;
RX   MEDLINE=21128732; PubMed=11234002;
RA   Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA   Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA   Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA   Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA   Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA   Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA   Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA   Barrell B.G.;
RT   "Massive gene decay in the leprosy bacillus.";
RL   Nature 409:1007-1011(2001).
CC   -!- FUNCTION: CLEAVAGE OF GUANOSINE OR INOSINE TO RESPECTIVE BASES AND
CC       SUGAR-1-PHOSPHATE MOLECULES (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: Purine nucleoside + phosphate = purine +
CC       alpha-D-ribose 1-phosphate.
CC   -!- PATHWAY: Purine nucleoside salvage.
CC   -!- SIMILARITY: BELONGS TO THE PNP/MTAP FAMILY 2 OF PHOSPHORYLASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U00022; AAA17341.1; -.
DR   EMBL; AL583919; CAC30216.1; -.
DR   PIR; S73042; S73042.
DR   HSSP; P81989; 1QE5.
DR   Leproma; ML0707; -.
DR   InterPro; IPR001369; Mtap_PNP.
DR   Pfam; PF00896; Mtap_PNP; 1.
DR   PROSITE; PS01240; PNP_MTAP_2; 1.
KW   Transferase; Glycosyltransferase; Complete proteome.
SQ   SEQUENCE   268 AA;  27980 MW;  46C622532FC96A0F CRC64;

  Query Match             78.3%;  Score 36;  DB 1;  Length 268;
  Best Local Similarity   75.0%;  Pred. No. 7.2;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||| ||
Db         95 HDLRHVVH 102


RESULT 6
BXE_CLOBO
ID   BXE_CLOBO      STANDARD;      PRT;  1250 AA.
AC   Q00496;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Beluga;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92174922; PubMed=1541280;
RA   Whelan S.M., Elmore M.J., Bodsworth N.J., Atkinson T., Minton N.P.;
RT   "The complete amino acid sequence of the Clostridium botulinum type-E
RT   neurotoxin, derived by nucleotide-sequence analysis of the encoding
RT   gene.";
RL   Eur. J. Biochem. 204:657-667(1992).
RN   [3]
RP   SEQUENCE OF 1-251 FROM N.A.
RX   MEDLINE=90264400; PubMed=2160960;
RA   Binz T., Kurazono H., Wille M., Frevert J., Wernars K., Niemann H.;
RT   "The complete sequence of botulinum neurotoxin type A and comparison
RT   with other clostridial neurotoxins.";
RL   J. Biol. Chem. 265:9153-9158(1990).
```

```
RN   [4]
RP   SEQUENCE OF 1-13.
RX   MEDLINE=85197963; PubMed=3888113;
RA   Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequences of botulinum neurotoxins types B and
RT   E.";
RL   Arch. Biochem. Biophys. 238:544-548(1985).
RN   [5]
RP   SEQUENCE OF 419-426.
RX   MEDLINE=90344918; PubMed=2116911;
RA   Gimenez J.A., Dasgupta B.R.;
RT   "Botulinum neurotoxin type E fragmented with endoproteinase Lys-C
RT   reveals the site trypsin nicks and homology with tetanus
RT   neurotoxin.";
RL   Biochimie 72:213-217(1990).
RN   [6]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 180-ARG-|-ILE-
CC       181 BOND IN SNAP-25.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC       HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC       WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC       FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC       BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62089; CAA43999.1; -.
DR   EMBL; X62683; CAA44558.1; -.
DR   PIR; S21178; S21178.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR006025; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    176    176       R -> G (IN REF. 2).
FT   CONFLICT    197    197       C -> S (IN REF. 2 AND 3).
FT   CONFLICT    339    339       R -> A (IN REF. 2).
FT   CONFLICT    772    772       I -> L (IN REF. 2).
FT   CONFLICT    962    963       FE -> LQ (IN REF. 2).
FT   CONFLICT    966    966       R -> A (IN REF. 2).
FT   CONFLICT   1194   1194       N -> NN (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143712 MW;  D9FCE26DDA041EB4 CRC64;

  Query Match             78.3%;  Score 36;  DB 1;  Length 1250;
  Best Local Similarity   75.0%;  Pred. No. 37;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||| ||
Db        211 HELIHSLH 218


RESULT 7
BXE_CLOBU
ID   BXE_CLOBU      STANDARD;      PRT;  1250 AA.
AC   P30995;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 43181, and ATCC 43755;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE OF 1-251 FROM N.A.
RC   STRAIN=BL6340;
RX   MEDLINE=91237316; PubMed=2033376;
RA   Fujii N., Kimura K., Murakami T., Indoh T., Tsuzuki K.,
RA   Yokosawa N., Yashiki T., Oguma K.;
RT   "Cloning of a DNA fragment encoding the 5'-terminus of the botulinum
RT   type E toxin gene from Clostridium butyricum strain BL6340.";
RL   J. Gen. Microbiol. 137:519-525(1991).
RN   [3]
RP   SEQUENCE OF 1-48.
RC   STRAIN=5262;
RA   Gimenez J., Foley J., Dasgupta B.R.;
RT   "Neurotoxin type E from Clostridium botulinum and C. butyricum;
RT   partial sequence and comparison.";
RL   FASEB J. 2:A1750-A1750(1988).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC       HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC       WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
```

```
:C       FORMATION AND TOXIN BINDING, RESPECTIVELY.
:C   -!- SUBCELLULAR LOCATION: Secreted.
:C   -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
:C       BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
:C   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
:C   ---------------------------------------------------------------------
:C   This SWISS-PROT entry is copyright. It is produced through a collaboration
:C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
:C   the European Bioinformatics Institute. There are no restrictions on its
:C   use by non-profit institutions as long as its content is in no way
:C   modified and this statement is not removed. Usage by and for commercial
:C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
:C   or send an email to license@isb-sib.ch).
:C   ---------------------------------------------------------------------
)R   EMBL; X62088; CAA43998.1; -.
)R   EMBL; X53180; CAA37321.1; -.
)R   PIR; JH0256; JH0256.
)R   HSSP; P10845; 3BTA.
)R   MEROPS; M27.002; -.
)R   InterPro; IPR000395; Bontoxilysin.
)R   InterPro; IPR006025; Zn_MTpeptdse.
)R   Pfam; PF01742; Peptidase_M27; 1.
)R   PRINTS; PR00760; BONTOXILYSIN.
)R   ProDom; PD001963; Bontoxilysin; 1.
)R   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    229    229       K -> M (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143265 MW;  8171B5B2C2312857 CRC64;

  Query Match             78.3%;  Score 36;  DB 1;  Length 1250;
  Best Local Similarity   75.0%;  Pred. No. 37;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||| ||
Db        211 HELIHSLH 218


RESULT 8.
RECR_LISIN
ID   RECR_LISIN     STANDARD;      PRT;   198 AA.
AC   Q927D9;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Recombination protein recR.
GN   RECR OR LIN2850.
OS   Listeria innocua.
OC   Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX   NCBI_TaxID=1642;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CLIP 11262 / Serovar 6a;
RX   MEDLINE=21537279; PubMed=11679669;
RA   Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA   Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA   Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA   Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA   Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA   Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA   Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA   Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA   Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA   Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA   Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT   "Comparative genomics of Listeria species.";
RL   Science 294:849-852(2001).
CC   -!- FUNCTION: May play a role in DNA repair. It seems to be involved
CC       in an recBC-independent recombinational process of DNA repair. It
CC       may act with recF and recO (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE RECR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL596173; CAC98076.1; -.
DR   PIR; AD1788; AD1788.
DR   ListiList; LIN02850; -.
DR   HAMAP; MF_00017; -; 1.
DR   InterPro; IPR003583; HHH_1.
DR   InterPro; IPR000093; RecR.
DR   InterPro; IPR006171; Toprim_dom.
DR   InterPro; IPR006154; Toprim_sub.
DR   Pfam; PF02132; RecR; 1.
DR   Pfam; PF01751; Toprim; 1.
DR   SMART; SM00278; HhH1; 1.
DR   SMART; SM00493; TOPRIM; 1.
DR   TIGRFAMs; TIGR00615; recR; 1.
DR   PROSITE; PS01300; RECR; 1.
KW   DNA repair; DNA recombination; Zinc-finger; Complete proteome.
FT   ZN_FING      57     72       C4-TYPE (POTENTIAL).
SQ   SEQUENCE   198 AA;  21996 MW;  8A82E1A16415DFEF CRC64;

  Query Match             73.9%;  Score 34;  DB 1;  Length 198;
  Best Local Similarity   75.0%;  Pred. No. 12;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              | | ||||
Db        101 HGLYHVLH 108


RESULT 9
RECR_LISMO
ID   RECR_LISMO     STANDARD;      PRT;   198 AA.
AC   Q8Y3X7;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Recombination protein recR.
GN   RECR OR LMO2702.
OS   Listeria monocytogenes.
OC   Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX   NCBI_TaxID=1639;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EGD-e / Serovar 1/2a;
RX   MEDLINE=21537279; PubMed=11679669;
RA   Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA   Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA   Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA   Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA   Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA   Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA   Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA   Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA   Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA   Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA   Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT   "Comparative genomics of Listeria species.";
RL   Science 294:849-852(2001).
CC   -!- FUNCTION: May play a role in DNA repair. It seems to be involved
CC       in an recBC-independent recombinational process of DNA repair. It
CC       may act with recF and recO (By similarity).
```

```
:C   -!- SIMILARITY: BELONGS TO THE RECR FAMILY.
:C   ---------------------------------------------------------------------
:C   This SWISS-PROT entry is copyright. It is produced through a collaboration
:C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
:C   the European Bioinformatics Institute. There are no restrictions on its
:C   use by non-profit institutions as long as its content is in no way
:C   modified and this statement is not removed. Usage by and for commercial
:C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
:C   or send an email to license@isb-sib.ch).
:C   ---------------------------------------------------------------------
)R   EMBL; AL591984; CAD00915.1; -.
)R   PIR; AE1412; AE1412.
)R   ListiList; LMO02702; -.
)R   HAMAP; MF_00017; -; 1.
)R   InterPro; IPR003583; HHH_1.
)R   InterPro; IPR000093; RecR.
)R   InterPro; IPR006171; Toprim_dom.
)R   InterPro; IPR006154; Toprim_sub.
)R   Pfam; PF02132; RecR; 1.
)R   Pfam; PF01751; Toprim; 1.
)R   SMART; SM00278; HhH1; 1.
)R   SMART; SM00493; TOPRIM; 1.
)R   TIGRFAMs; TIGR00615; recR; 1.
)R   PROSITE; PS01300; RECR; 1.
(W   DNA repair; DNA recombination; Zinc-finger; Complete proteome.
?T   ZN_FING      57      72       C4-TYPE (POTENTIAL).
3Q   SEQUENCE   198 AA;  21934 MW;  E542E27BC3D05036 CRC64;

  Query Match             73.9%;  Score 34;  DB 1;  Length 198;
  Best Local Similarity   75.0%;  Pred. No. 12;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

)y          1 HDLIHVLH 8
              | | ||||
)b        101 HGLYHVLH 108


RESULT 10
RECR_STRPN
ID   RECR_STRPN     STANDARD;      PRT;   198 AA.
AC   Q9ZHC4;
)T   30-MAY-2000 (Rel. 39, Created)
)T   30-MAY-2000 (Rel. 39, Last sequence update)
)T   28-FEB-2003 (Rel. 41, Last annotation update)
)E   Recombination protein recR.
GN   RECR OR RECM OR SP1672 OR SPR1516.
)S   Streptococcus pneumoniae, and
)S   Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
)C   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
)C   Streptococcus.
)X   NCBI_TaxID=1313, 171101;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=G54 / Type 19F;
RX   MEDLINE=99061199; PubMed=9846742;
RA   Massidda O., Anderluzzi D., Friedli L., Feger G.;
RT   "Unconventional organization of the division and cell wall gene
RT   cluster of Streptococcus pneumoniae.";
RL   Microbiology 144:3069-3078(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=339, and PN94-661;
RX   MEDLINE=20073037; PubMed=10605111;
RA   Enright M.C., Spratt B.G.;
RT   "Extensive variation in the ddl gene of penicillin-resistant
RT   Streptococcus pneumoniae results from a hitchhiking effect driven by
RT   the penicillin-binding protein 2b gene.";
RL   Mol. Biol. Evol. 16:1687-1695(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC BAA-334 / TIGR4;
RX   MEDLINE=21357209; PubMed=11463916;
RA   Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA   Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA   Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA   Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA   Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA   McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA   Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA   Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT   "Complete genome sequence of a virulent isolate of Streptococcus
RT   pneumoniae.";
RL   Science 293:498-506(2001).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC BAA-255 / R6;
RX   MEDLINE=21429245; PubMed=11544234;
RA   Hoskins J., Alborn W.E. Jr., Arnold J., Blaszczak L.C., Burgett S.,
RA   DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA   Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA   LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA   McAhren S.M., McHenney M., McLeaster K., Mundy C.W., Nicas T.I.,
RA   Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
RA   Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA   Zook C.A., Baltz R.H., Jaskunas S.R., Rosteck P.R. Jr., Skatrud P.L.,
RA   Glass J.I.;
RT   "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL   J. Bacteriol. 183:5709-5717(2001).
CC   -!- FUNCTION: MAY PLAY A ROLE IN DNA REPAIR. IT SEEMS TO BE INVOLVED
CC       IN AN RECBC-INDEPENDENT RECOMBINATIONAL PROCESS OF DNA REPAIR. IT
CC       MAY ACT WITH RECF AND RECO (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE RECR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF068901; AAC95434.1; -.
DR   EMBL; AJ243056; CAB64474.1; -.
DR   EMBL; AJ243057; CAB64478.1; -.
DR   EMBL; AE007460; AAK75751.1; -.
DR   EMBL; AE008520; AAL00320.1; -.
DR   PIR; C98061; C98061.
DR   PIR; F95194; F95194.
DR   TIGR; SP1672; -.
DR   HAMAP; MF_00017; -; 1.
DR   InterPro; IPR000093; RecR.
DR   InterPro; IPR006171; Toprim_dom.
DR   InterPro; IPR006154; Toprim_sub.
DR   Pfam; PF02132; RecR; 1.
DR   Pfam; PF01751; Toprim; 1.
DR   SMART; SM00493; TOPRIM; 1.
DR   TIGRFAMs; TIGR00615; recR; 1.
DR   PROSITE; PS01300; RECR; 1.
KW   DNA repair; DNA recombination; Zinc-finger; Complete proteome.
FT   ZN_FING      57      72       C4-TYPE (POTENTIAL).
SQ   SEQUENCE   198 AA;  21689 MW;  FC6F0E98E3933752 CRC64;

  Query Match             73.9%;  Score 34;  DB 1;  Length 198;
  Best Local Similarity   75.0%;  Pred. No. 12;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              | | ||||
Db        101 HGLYHVLH 108


RESULT 11
SODC_HAEDU
ID   SODC_HAEDU     STANDARD;      PRT;   199 AA.
AC   Q59452; Q59449; Q59453;
DT   15-DEC-1998 (Rel. 37, Created)
```

```
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Superoxide dismutase [Cu-Zn] precursor (EC 1.15.1.1).
GN   SODC.
OS   Haemophilus ducreyi.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC   Pasteurellaceae; Haemophilus.
OX   NCBI_TaxID=730;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=35000;
RX   MEDLINE=97288949; PubMed=9143881;
RA   Langford P.R., Kroll J.S.;
RT   "Distribution, cloning, characterisation and mutagenesis of sodC, the
RT   gene encoding copper/zinc superoxide dismutase, a potential
RT   determinant of virulence, in Haemophilus ducreyi.";
RL   FEMS Immunol. Med. Microbiol. 17:235-242(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=35000;
RX   MEDLINE=97149276; PubMed=8996084;
RA   Stevens M.K., Hassett D.J., Radolf J.D., Hansen E.J.;
RT   "Cloning and sequencing of the gene encoding the Cu,Zn-superoxide
RT   dismutase of Haemophilus ducreyi.";
RL   Gene 183:35-40(1996).
RN   [3]
RP   SEQUENCE OF 100-186 FROM N.A.
RC   STRAIN=35000;
RX   MEDLINE=96118708; PubMed=7496539;
RA   Kroll J.S., Langford P.R., Wilks K.E., Keil A.D.;
RT   "Bacterial [Cu,Zn]-superoxide dismutase: phylogenetically distinct
RT   from the eukaryotic enzyme, and not so rare after all!";
RL   Microbiology 141:2271-2279(1995).
CC   -!- FUNCTION: Destroys radicals which are normally produced within the
CC       cells and which are toxic to biological systems. May play a role
CC       in the interactive biology of organisms with their hosts and so
CC       contribute to their capacity to cause disease.
CC   -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC   -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
CC       similarity).
CC   -!- SUBUNIT: Homodimer (By similarity).
CC   -!- SUBCELLULAR LOCATION: Periplasmic.
CC   -!- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X98737; CAA67289.1; -.
DR   EMBL; U47664; AAB41293.1; -.
DR   EMBL; X83125; CAA58206.1; -.
DR   PIR; JC5718; JC5718.
DR   HSSP; P24702; 2APS.
DR   InterPro; IPR001424; SOD_CU_ZN.
DR   Pfam; PF00080; sodcu; 1.
DR   ProDom; PD000469; SOD_CU_ZN; 1.
DR   PROSITE; PS00087; SOD_CU_ZN_1; 1.
DR   PROSITE; PS00332; SOD_CU_ZN_2; 1.
KW   Antioxidant; Oxidoreductase; Metal-binding; Copper; Zinc; Periplasmic;
KW   Signal.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    199       SUPEROXIDE DISMUTASE [CU-ZN].
FT   METAL        92     92       COPPER (BY SIMILARITY).
FT   METAL        94     94       COPPER (BY SIMILARITY).
FT   METAL       117    117       COPPER AND ZINC (BY SIMILARITY).
FT   METAL       126    126       ZINC (BY SIMILARITY).
FT   METAL       135    135       ZINC (BY SIMILARITY).
FT   METAL       138    138       ZINC (BY SIMILARITY).
FT   METAL       173    173       COPPER (BY SIMILARITY).
FT   DISULFID     99    195       BY SIMILARITY.
SQ   SEQUENCE   199 AA;  21402 MW;  841D3210AB2BC06C CRC64;

  Query Match             73.9%;  Score 34;  DB 1;  Length 199;
  Best Local Similarity   75.0%;  Pred. No. 13;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              ||| | ||
Db         82 HDLAHGLH 89


RESULT 12
EX7L_VIBCH
ID   EX7L_VIBCH     STANDARD;      PRT;   446 AA.
AC   Q9KTW4;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Probable exodeoxyribonuclease VII large subunit (EC 3.1.11.6)
DE   (Exonuclease VII large subunit).
GN   XSEA OR VC0766.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC   Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=El Tor N16961 / Serotype O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
CC   -!- FUNCTION: BIDIRECTIONALLY DEGRADES SINGLE-STRANDED DNA INTO LARGE
CC       ACID-INSOLUBLE OLIGONUCLEOTIDES, WHICH ARE THEN DEGRADED FURTHER
CC       INTO SMALL ACID-SOLUBLE OLIGONUCLEOTIDES (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage in either 5'- to 3'-
CC       or 3'- to 5'-direction to yield nucleoside 5'-phosphates.
CC   -!- SUBUNIT: Heterooligomer composed of large and small subunits (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE XSEA FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE004162; AAF93931.1; -.
DR   PIR; B82282; B82282.
DR   TIGR; VC0766; -.
DR   HAMAP; MF_00378; -; 1.
DR   InterPro; IPR003753; Exonuc_VII_L.
DR   InterPro; IPR004365; tRNA_anti.
DR   Pfam; PF02601; Exonuc_VII_L; 1.
DR   Pfam; PF01336; tRNA_anti; 1.
DR   TIGRFAMs; TIGR00237; xseA; 1.
KW   Hydrolase; Nuclease; Exonuclease; Complete proteome.
SQ   SEQUENCE   446 AA;  50542 MW;  AAE17369636A4BC9 CRC64;

  Query Match             73.9%;  Score 34;  DB 1;  Length 446;
  Best Local Similarity   71.4%;  Pred. No. 30;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 HDLIHVL 7
              ||::|||
Db        155 HDILHVL 161


RESULT 13
EX7L_SALTI
ID   EX7L_SALTI     STANDARD;      PRT;   449 AA.
AC   Q8Z4Q1;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Probable exodeoxyribonuclease VII large subunit (EC 3.1.11.6)
DE   (Exonuclease VII large subunit).
GN   XSEA OR STY2753 OR T0345.
OS   Salmonella typhi.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Salmonella.
OX   NCBI_TaxID=601;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CT18;
RX   MEDLINE=21534947; PubMed=11677608;
RA   Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA   Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA   Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA   Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA   Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA   Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA   Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA   Whitehead S., Barrell B.G.;
RT   "Complete genome sequence of a multiple drug resistant Salmonella
RT   enterica serovar Typhi CT18.";
RL   Nature 413:848-852(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Ty2 / ATCC 700931;
RX   MEDLINE=22531367; PubMed=12644504;
RA   Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA   Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RT   "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT   and CT18.";
RL   J. Bacteriol. 185:2330-2337(2003).
CC   -!- FUNCTION: Bidirectionally degrades single-stranded DNA into large
CC       acid-insoluble oligonucleotides, which are then degraded further
CC       into small acid-soluble oligonucleotides (By similarity).
CC   -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage in either 5'- to 3'-
CC       or 3'- to 5'-direction to yield nucleoside 5'-phosphates.
CC   -!- SUBUNIT: Heterooligomer composed of large and small subunits (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE XSEA FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL627275; CAD02714.1; -.
DR   EMBL; AE016835; AAO68065.1; -.
DR   HAMAP; MF_00378; -; 1.
DR   InterPro; IPR003753; Exonuc_VII_L.
DR   InterPro; IPR004365; tRNA_anti.
DR   Pfam; PF02601; Exonuc_VII_L; 1.
DR   Pfam; PF01336; tRNA_anti; 1.
DR   TIGRFAMs; TIGR00237; xseA; 1.
KW   Hydrolase; Nuclease; Exonuclease; Complete proteome.
SQ   SEQUENCE   449 AA;  50720 MW;  511957DEC878F5D2 CRC64;

  Query Match             73.9%;  Score 34;  DB 1;  Length 449;
  Best Local Similarity   71.4%;  Pred. No. 30;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HDLIHVL 7
              ||::|||
Db        154 HDILHVL 160


RESULT 14
EX7L_SALTY
ID   EX7L_SALTY     STANDARD;      PRT;   449 AA.
AC   Q8ZN58;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Probable exodeoxyribonuclease VII large subunit (EC 3.1.11.6)
DE   (Exonuclease VII large subunit).
GN   XSEA OR STM2512.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
CC   -!- FUNCTION: Bidirectionally degrades single-stranded DNA into large
CC       acid-insoluble oligonucleotides, which are then degraded further
CC       into small acid-soluble oligonucleotides (By similarity).
CC   -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage in either 5'- to 3'-
CC       or 3'- to 5'-direction to yield nucleoside 5'-phosphates.
CC   -!- SUBUNIT: Heterooligomer composed of large and small subunits (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE XSEA FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE008813; AAL21406.1; -.
DR   StyGene; SG?????; xseA.
DR   HAMAP; MF_00378; -; 1.
DR   InterPro; IPR003753; Exonuc_VII_L.
DR   InterPro; IPR004365; tRNA_anti.
DR   Pfam; PF02601; Exonuc_VII_L; 1.
DR   Pfam; PF01336; tRNA_anti; 1.
DR   TIGRFAMs; TIGR00237; xseA; 1.
KW   Hydrolase; Nuclease; Exonuclease; Complete proteome.
SQ   SEQUENCE   449 AA;  50613 MW;  85356CE8560E161E CRC64;

  Query Match             73.9%;  Score 34;  DB 1;  Length 449;
  Best Local Similarity   71.4%;  Pred. No. 30;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HDLIHVL 7
              ||::|||
Db        154 HDILHVL 160
```

```
ESULT 15
X7L_ECO57
D   EX7L_ECO57      STANDARD;      PRT;   456 AA.
C   Q8XAB0;
T   28-FEB-2003 (Rel. 41, Created)
T   28-FEB-2003 (Rel. 41, Last sequence update)
T   28-FEB-2003 (Rel. 41, Last annotation update)
E   Probable exodeoxyribonuclease VII large subunit (EC 3.1.11.6)
E   (Exonuclease VII large subunit).
N   XSEA OR Z3773 OR ECS3371.
S   Escherichia coli O157:H7.
C   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
C   Enterobacteriaceae; Escherichia.
X   NCBI_TaxID=83334;
N   [1]
P   SEQUENCE FROM N.A.
C   STRAIN=O157:H7 / EDL933 / ATCC 700927;
X   MEDLINE=21074935; PubMed=11206551;
A   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
A   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
A   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
A   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
A   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
A   Welch R.A., Blattner F.R.;
T   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
L   Nature 409:529-533(2001).
N   [2]
P   SEQUENCE FROM N.A.
C   STRAIN=O157:H7 / RIMD 0509952;
X   MEDLINE=21156231; PubMed=11258796;
A   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
A   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
A   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
A   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
T   "Complete genome sequence of enterohemorrhagic Escherichia coli
T   O157:H7 and genomic comparison with a laboratory strain K-12.";
L   DNA Res. 8:11-22(2001).
C   -!- FUNCTION: Bidirectionally degrades single-stranded DNA into large
C       acid-insoluble oligonucleotides, which are then degraded further
C       into small acid-soluble oligonucleotides (By similarity).
C   -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage in either 5'- to 3'-
C       or 3'- to 5'-direction to yield nucleoside 5'-phosphates.
C   -!- SUBUNIT: Heterooligomer composed of large and small subunits (By
C       similarity).
C   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
C   -!- SIMILARITY: BELONGS TO THE XSEA FAMILY.
C   ----------------------------------------------------------------------
C   This SWISS-PROT entry is copyright. It is produced through a collaboration
C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
C   the European Bioinformatics Institute. There are no restrictions on its
C   use by non-profit institutions as long as its content is in no way
C   modified and this statement is not removed. Usage by and for commercial
C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
C   or send an email to license@isb-sib.ch).
C   ----------------------------------------------------------------------
R   EMBL; AE005480; AAG57620.1; -.
R   EMBL; AP002561; BAB36794.1; -.
R   PIR; C91050; C91050.
R   HAMAP; MF_00378; -; 1.
R   InterPro; IPR003753; Exonuc_VII_L.
R   InterPro; IPR004365; tRNA_anti.
R   Pfam; PF02601; Exonuc_VII_L; 1.
R   Pfam; PF01336; tRNA_anti; 1.
R   TIGRFAMs; TIGR00237; xseA; 1.
W   Hydrolase; Nuclease; Exonuclease; Complete proteome.
Q   SEQUENCE   456 AA;  51734 MW;  174EAE7F72EB3C37 CRC64;

 Query Match             73.9%;  Score 34;  DB 1;  Length 456;
 Best Local Similarity   71.4%;  Pred. No. 31;
 Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HDLIHVL 7
              ||::|||
Db        154 HDILHVL 160

Search completed: November 21, 2003, 15:48:34
Job time : 11 secs
```

GenCore version 5.1.6


)M protein - protein search, using sw model

:un on: November 21, 2003, 15:43:40 ; Search time 28 Seconds
(without alignments)
73.729 Million cell updates/sec

'itle: US-10-064-903-2
'erfect score: 46
;equence: 1 HDLIHVLH 8

:coring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

;earched: 830525 seqs, 258052604 residues

'otal number of hits satisfying chosen parameters: 830525

'inimum DB seq length: 0
'aximum DB seq length: 2000000000

'ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

)atabase : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 42 | 91.3 | 451 | 2 | Q9R631 | Q9r631 clostridium |
| 2 | 42 | 91.3 | 1268 | 2 | Q45851 | Q45851 clostridium |
| 3 | 42 | 91.3 | 1291 | 2 | Q9ZAJ8 | Q9zaj8 clostridium |
| 4 | 42 | 91.3 | 1291 | 2 | Q93G71 | Q93g71 clostridium |
| 5 | 42 | 91.3 | 1291 | 2 | Q933K0 | Q933k0 clostridium |
| 6 | 42 | 91.3 | 1291 | 2 | Q08077 | Q08077 clostridium |
| 7 | 42 | 91.3 | 1291 | 2 | Q8GR96 | Q8gr96 clostridium |
| 8 | 42 | 91.3 | 1310 | 2 | Q93N27 | Q93n27 clostridium |
| 9 | 39 | 84.8 | 397 | 16 | Q92ML7 | Q92ml7 rhizobium m |
| 10 | 38 | 82.6 | 925 | 16 | Q9JS16 | Q9js16 chlamydia p |
| 11 | 38 | 82.6 | 925 | 16 | Q9Z8P5 | Q9z8p5 chlamydia p |
| 12 | 38 | 82.6 | 1278 | 2 | Q57236 | Q57236 clostridium |
| 13 | 38 | 82.6 | 1280 | 2 | Q9ZAJ5 | Q9zaj5 clostridium |
| 14 | 37 | 80.4 | 349 | 5 | Q8IJV0 | Q8ijv0 plasmodium |
| 15 | 37 | 80.4 | 707 | 3 | Q8X008 | Q8x008 neurospora |
| 16 | 36 | 78.3 | 105 | 10 | Q8VXL7 | Q8vxl7 fagus sylva |
| 17 | 36 | 78.3 | 232 | 16 | Q8X829 | Q8x829 escherichia |
| 18 | 36 | 78.3 | 237 | 13 | Q8AWC9 | Q8awc9 cyprinus ca |
| 19 | 36 | 78.3 | 241 | 10 | Q8VXL6 | Q8vxl6 fagus sylva |
| 20 | 36 | 78.3 | 489 | 9 | Q9AZH2 | Q9azh2 bacteriopha |
| 21 | 36 | 78.3 | 489 | 16 | Q9CEA2 | Q9cea2 lactococcus |
| 22 | 36 | 78.3 | 773 | 10 | Q9SHP2 | Q9shp2 arabidopsis |
| 23 | 36 | 78.3 | 1251 | 2 | Q9K395 | Q9k395 clostridium |
| 24 | 36 | 78.3 | 1252 | 2 | Q8KZM3 | Q8kzm3 clostridium |
| 25 | 36 | 78.3 | 1255 | 2 | Q9FAR6 | Q9far6 clostridium |
| 26 | 35 | 76.1 | 312 | 16 | O26068 | O26068 helicobacte |
| 27 | 35 | 76.1 | 312 | 16 | Q9ZJ59 | Q9zj59 helicobacte |
| 28 | 35 | 76.1 | 426 | 5 | Q8MYP4 | Q8myp4 caenorhabdi |
| 29 | 35 | 76.1 | 431 | 5 | Q9XTZ9 | Q9xtz9 caenorhabdi |
| 30 | 35 | 76.1 | 500 | 16 | Q9RVQ8 | Q9rvq8 deinococcus |
| 31 | 35 | 76.1 | 679 | 16 | Q97SL8 | Q97sl8 streptococc |
| 32 | 35 | 76.1 | 737 | 16 | Q8DR83 | Q8dr83 streptococc |
| 33 | 34 | 73.9 | 129 | 3 | Q8TFT7 | Q8tft7 ustilago vi |
| 34 | 34 | 73.9 | 198 | 2 | Q9RCP8 | Q9rcp8 streptococc |
| 35 | 34 | 73.9 | 198 | 2 | Q9RCQ5 | Q9rcq5 streptococc |
| 36 | 34 | 73.9 | 198 | 2 | Q9RCR0 | Q9rcr0 streptococc |
| 37 | 34 | 73.9 | 198 | 2 | Q9R2M1 | Q9r2m1 streptococc |
| 38 | 34 | 73.9 | 198 | 2 | Q9RCQ2 | Q9rcq2 streptococc |
| 39 | 34 | 73.9 | 199 | 16 | Q8DV99 | Q8dv99 streptococc |
| 40 | 34 | 73.9 | 222 | 16 | Q8EE87 | Q8ee87 shewanella |
| 41 | 34 | 73.9 | 225 | 5 | Q8I1V8 | Q8i1v8 plasmodium |
| 42 | 34 | 73.9 | 261 | 5 | Q8IFQ1 | Q8ifq1 plasmodium |
| 43 | 34 | 73.9 | 299 | 17 | Q96YN4 | Q96yn4 sulfolobus |
| 44 | 34 | 73.9 | 385 | 10 | Q8H3N0 | Q8h3n0 oryza sativ |
| 45 | 34 | 73.9 | 458 | 16 | Q8FF64 | Q8ff64 escherichia |

ALIGNMENTS

RESULT 1
Q9R631
ID Q9R631 PRELIMINARY; PRT; 451 AA.
AC Q9R631;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Neurotoxin type B light chain, BONT/B.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92340509; PubMed=1634516;
RA Kurazono H., Mochida S., Binz T., Eisel U., Quanz M., Grebenstein O.,
RA Wernars K., Poulain B., Tauc L., Niemann H.;
RT "Minimal essential domains specifying toxicity of the light chains of
RT tetanus toxin and botulinum neurotoxin type A.";
RL J. Biol. Chem. 267:14721-14729(1992).
DR HSSP; P10845; 3BTA.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn_MTpeptdse.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 451 AA; 51943 MW; 6C79FD488653EA71 CRC64;

Query Match 91.3%; Score 42; DB 2; Length 451;
Best Local Similarity 87.5%; Pred. No. 6.3;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 HDLIHVLH 8
              |:||||||
Db        230 HELIHVLH 237
```

RESULT 2

```
)45851
'D   Q45851      PRELIMINARY;      PRT;  1268 AA.
AC   Q45851;
)T   01-NOV-1996 (TrEMBLrel. 01, Created)
)T   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
)T   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
)E   Neurotoxin type F.
)N   BONT /F.
)S   Clostridium baratii.
)C   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
)C   Clostridium.
)X   NCBI_TaxID=1561;
)N   [1]
)P   SEQUENCE FROM N.A.
)X   MEDLINE=93252228; PubMed=8486245;
)A   Thompson D.E., Hutson R.A., East A.K., Allaway D., Collins M.D.,
)A   Richardson P.T.;
)T   "Nucleotide sequence of the gene coding for Clostridium barati type F
)T   neurotoxin: Comparison with other clostridial neurotoxins.";
)L   FEMS Microbiol. Lett. 108:175-182(1993).
)R   EMBL; X68262; CAA48329.1; -.
)R   HSSP; P10845; 3BTA.
)R   MEROPS; M27.002; -.
)R   InterPro; IPR000395; Bontoxilysin.
)R   InterPro; IPR006025; Zn_MTpeptdse.
)R   Pfam; PF01742; Peptidase_M27; 1.
)R   ProDom; PD001963; Bontoxilysin; 1.
)R   PROSITE; PS00142; ZINC_PROTEASE; 1.
)Q   SEQUENCE   1268 AA;  145513 MW;  963040091AC15ED2 CRC64;

  Query Match             91.3%;  Score 42;  DB 2;  Length 1268;
  Best Local Similarity   87.5%;  Pred. No. 18;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

)y          1 HDLIHVLH 8
              |:||||||
)b        219 HELIHVLH 226


RESULT 3
)9ZAJ8
ID   Q9ZAJ8      PRELIMINARY;      PRT;  1291 AA.
AC   Q9ZAJ8;
)T   01-MAY-1999 (TrEMBLrel. 10, Created)
)T   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
)T   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
)E   BonT protein.
GN   BONT.
)S   Clostridium botulinum.
)C   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
)C   Clostridium.
)X   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 3281;
RX   MEDLINE=98440323; PubMed=9767710;
RA   Santos-Buelga J., Collins M.D., East A.K.;
RT   "Characterization of the genes encoding the Botulinum neurotoxin
RT   complex in a strain of clostridium botulinum producing type B & F
RT   neurotoxins.";
RL   Curr. Microbiol. 37:312-318(1998).
DR   EMBL; Y13630; CAA73968.1; -.
DR   HSSP; P10845; 3BTA.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR006025; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150840 MW;  E4D3B0E46AB2E735 CRC64;

  Query Match             91.3%;  Score 42;  DB 2;  Length 1291;
  Best Local Similarity   87.5%;  Pred. No. 18;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||||||
Db        230 HELIHVLH 237


RESULT 4
Q93G71
ID   Q93G71      PRELIMINARY;      PRT;  1291 AA.
AC   Q93G71;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Neurotoxin type B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1436;
RA   Kirma N., Ferreira J.L., Baumstark B.R.;
RT   "Characterization of six type A strains of Clostridium botulinum that
RT   contain type B toxin gene sequences.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF295926; AAK97132.1; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR006025; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150824 MW;  D7CA07BAE2EB8CD2 CRC64;

  Query Match             91.3%;  Score 42;  DB 2;  Length 1291;
  Best Local Similarity   87.5%;  Pred. No. 18;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||||||
Db        230 HELIHVLH 237


RESULT 5
Q933K0
ID   Q933K0      PRELIMINARY;      PRT;  1291 AA.
AC   Q933K0;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Type B cryptic neurotoxin.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=593, and 588;
RA   Kirma N., Ferreira J.L., Baumstark B.R.;
RT   "Characterization of six type A strains of Clostridium botulinum that
RT   contain type B toxin gene sequences.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF300466; AAL11499.1; -.
DR   EMBL; AF300465; AAL11498.1; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR006025; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1291 AA;  150843 MW;  7AC1737B0FA5A151 CRC64;
```

```
  Query Match             91.3%;  Score 42;  DB 2;  Length 1291;
  Best Local Similarity   87.5%;  Pred. No. 18;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

)y          1 HDLIHVLH 8
              |:||||||
)b        230 HELIHVLH 237


:ESULT 6
)08077
:D   Q08077      PRELIMINARY;      PRT;  1291 AA.
\C   Q08077;
)T   01-NOV-1996 (TrEMBLrel. 01, Created)
)T   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
)T   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
)E   BoNT/B.
)N   BONT/B.
)S   Clostridium botulinum.
)C   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
)C   Clostridium.
)X   NCBI_TaxID=1491;
:N   [1]
:P   SEQUENCE FROM N.A.
:C   STRAIN=Eklund 17B ATCC25765;
:X   MEDLINE=94122659; PubMed=7764370;
:A   Hutson R.A., Collins M.D., East A.K., Thompson D.E.;
:T   "Nucleotide sequence of the gene coding for non-proteolytic
:T   Clostridium botulinum type B neurotoxin: comparison with other
:T   clostridial neurotoxins.";
:L   Curr. Microbiol. 28:101-110(1994).
)R   EMBL; X71343; CAA50482.1; -.
)R   HSSP; P10845; 3BTA.
)R   MEROPS; M27.002; -.
)R   InterPro; IPR000395; Bontoxilysin.
)R   InterPro; IPR006025; Zn_MTpeptdse.
)R   Pfam; PF01742; Peptidase_M27; 1.
)R   PRINTS; PR00760; BONTOXILYSIN.
)R   ProDom; PD001963; Bontoxilysin; 1.
)R   PROSITE; PS00142; ZINC_PROTEASE; 1.
:Q   SEQUENCE   1291 AA;  150513 MW;  71BCAFE23D69FAAA CRC64;

  Query Match             91.3%;  Score 42;  DB 2;  Length 1291;
  Best Local Similarity   87.5%;  Pred. No. 18;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

)y          1 HDLIHVLH 8
              |:||||||
)b        230 HELIHVLH 237


:ESULT 7
)8GR96
:D   Q8GR96      PRELIMINARY;      PRT;  1291 AA.
\C   Q8GR96;
)T   01-MAR-2003 (TrEMBLrel. 23, Created)
)T   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
)T   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
)E   Neurotoxin.
:N   BONTB.
)S   Clostridium botulinum.
)C   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
)C   Clostridium.
)X   NCBI_TaxID=1491;
:N   [1]
:P   SEQUENCE FROM N.A.
:A   Ihara H., Kohda T., Morimoto F., Tsukamoto K., Karasawa T.,
:A   Nakamura S., Mukamoto M., Kozaki S.;
:T   "Clostridium botulinum type B neurotoxin associated with infant
:T   botulism.";
:L   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
)R   EMBL; AB084152; BAC22064.1; -.
:Q   SEQUENCE   1291 AA;  150574 MW;  0227CAEF4F58504D CRC64;
```

```
  Query Match             91.3%;  Score 42;  DB 2;  Length 1291;
  Best Local Similarity   87.5%;  Pred. No. 18;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||||||
Db        230 HELIHVLH 237


RESULT 8
Q93N27
ID   Q93N27      PRELIMINARY;      PRT;  1310 AA.
AC   Q93N27;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Tetanus toxin (Fragment).
OS   Clostridium tetani.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1513;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Shumin Z., Dianliang L.;
RT   "Cloning and sequence analysis of tetanus toxin gene.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF389424; AAK72964.2; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR001064; Crystallin.
DR   InterPro; IPR006025; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
FT   NON_TER       1      1
FT   NON_TER    1310   1310
SQ   SEQUENCE   1310 AA;  150316 MW;  9EADDC914418E450 CRC64;

  Query Match             91.3%;  Score 42;  DB 2;  Length 1310;
  Best Local Similarity   87.5%;  Pred. No. 18;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||||||
Db        234 HELIHVLH 241


RESULT 9
Q92ML7
ID   Q92ML7      PRELIMINARY;      PRT;   397 AA.
AC   Q92ML7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Putative deaminase OR deamidase protein.
GN   R02596 OR SMC02420.
OS   Rhizobium meliloti (Sinorhizobium meliloti).
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Rhizobiaceae; Sinorhizobium.
OX   NCBI_TaxID=382;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1021;
RX   MEDLINE=21396507; PubMed=11481430;
RA   Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA   Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA   Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA   Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA   Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT   "Analysis of the chromosome sequence of the legume symbiont
RT   Sinorhizobium meliloti strain 1021.";
```

```
:L   Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
)R   EMBL; AL591791; CAC47175.1; -.
)R   InterPro; IPR006680; Amidohydro_1.
)R   InterPro; IPR001130; TatD_DNase.
)R   Pfam; PF01979; Amidohydro_1; 1.
)R   PROSITE; PS01137; TATD_1; 1.
;W   Complete proteome.
;Q   SEQUENCE   397 AA;  43054 MW;  B7D5F69C499CBE02 CRC64;

  Query Match             84.8%;  Score 39;  DB 16;  Length 397;
  Best Local Similarity   87.5%;  Pred. No. 20;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

)y          1 HDLIHVLH 8
              ||| ||||
)b        326 HDLEHVLH 333


(ESULT 10
)9JS16
'D   Q9JS16      PRELIMINARY;      PRT;   925 AA.
AC   Q9JS16;
)T   01-OCT-2000 (TrEMBLrel. 15, Created)
)T   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
)T   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
)E   CT234 hypothetical protein.
)N   CPJ0293 OR CP0465.
)S   Chlamydia pneumoniae (Chlamydophila pneumoniae).
)C   Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydophila.
)X   NCBI_TaxID=83558;
(N   [1]
(P   SEQUENCE FROM N.A.
(C   STRAIN=AR39;
(X   MEDLINE=20150255; PubMed=10684935;
(A   Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
(A   White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
(A   Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
(A   Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
(A   Eisen J., Fraser C.M.;
(T   "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
(T   pneumoniae AR39.";
(L   Nucleic Acids Res. 28:1397-1406(2000).
(N   [2]
(P   SEQUENCE FROM N.A.
(C   STRAIN=J138;
(X   MEDLINE=20330349; PubMed=10871362;
(A   Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
(A   Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
(T   "Comparison of whole genome sequences of Chlamydia pneumoniae J138
(T   from Japan and CWL029 from USA.";
(L   Nucleic Acids Res. 28:2311-2314(2000).
)R   EMBL; AE002208; AAF38302.1; -.
)R   EMBL; AP002546; BAA98503.1; -.
)R   TIGR; CP0465; -.
;Q   SEQUENCE   925 AA;  105601 MW;  61E8941E7C8FD620 CRC64;

  Query Match             82.6%;  Score 38;  DB 16;  Length 925;
  Best Local Similarity   62.5%;  Pred. No. 73;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

)y          1 HDLIHVLH 8
              |||:|: |
)b        538 HDLLHITH 545


(ESULT 11
)9Z8P5
ID   Q9Z8P5      PRELIMINARY;      PRT;   925 AA.
AC   Q9Z8P5;
)T   01-MAY-1999 (TrEMBLrel. 10, Created)
)T   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
)T   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
)E   CT234 hypothetical protein.
GN   CPN0293.
OS   Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC   Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX   NCBI_TaxID=83558;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CWL029;
RX   MEDLINE=99206606; PubMed=10192388;
RA   Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA   Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT   "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL   Nat. Genet. 21:385-389(1999).
DR   EMBL; AE001614; AAD18442.1; -.
KW   Complete proteome.
SQ   SEQUENCE   925 AA;  105615 MW;  98E6098E7C8FD37D CRC64;

  Query Match             82.6%;  Score 38;  DB 16;  Length 925;
  Best Local Similarity   62.5%;  Pred. No. 73;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |||:|: |
Db        538 HDLLHITH 545


RESULT 12
Q57236
ID   Q57236      PRELIMINARY;      PRT;  1278 AA.
AC   Q57236; Q45863;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Botulinum neurotoxin type F (BONT/F protein).
GN   BONT/F.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 10281;
RA   Hutson R.A., Collins M.D.;
RL   Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Elmore M.J., Bodsworth N.J., Whelan S.M., Minton N.P.;
RL   Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 635-1000 FROM N.A.
RC   STRAIN=NCTC 1028;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   SEQUENCE OF 1-27 FROM N.A.
RC   STRAIN=LANGELAND;
RX   MEDLINE=98404102; PubMed=9732534;
RA   East A.K., Bhandari M., Hielm S., Collins M.D.;
RT   "Analysis of the botulinum neurotoxin type F gene clusters in
RT   proteolytic and nonproteolytic Clostridium botulinum and Clostridium
RT   barati.";
RL   Curr. Microbiol. 37:262-268(1998).
DR   EMBL; X81714; CAA57358.1; -.
DR   EMBL; L35496; AAA23210.1; -.
DR   EMBL; X70821; CAA50152.1; -.
DR   EMBL; X99064; CAA67512.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR006025; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
```

```
)R   PRINTS; PR00760; BONTOXILYSIN.
)R   ProDom; PD001963; Bontoxilysin; 1.
)R   PROSITE; PS00142; ZINC_PROTEASE; 1.
CW   Neurotoxin.
3Q   SEQUENCE   1278 AA;  147073 MW;  A1BE1318431D6918 CRC64;

  Query Match             82.6%;  Score 38;  DB 2;  Length 1278;
  Best Local Similarity   75.0%;  Pred. No. 1e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||| ||
Db        227 HELIHALH 234


RESULT 13
Q9ZAJ5
ID   Q9ZAJ5      PRELIMINARY;      PRT;  1280 AA.
AC   Q9ZAJ5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   BonT protein.
GN   BONT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 3281;
RX   MEDLINE=98440323; PubMed=9767710;
RA   Santos-Buelga J., Collins M.D., East A.K.;
RT   "Characterization of the genes encoding the Botulinum neurotoxin
RT   complex in a strain of clostridium botulinum producing type B & F
RT   neurotoxins.";
RL   Curr. Microbiol. 37:312-318(1998).
DR   EMBL; Y13631; CAA73972.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR006025; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1280 AA;  147487 MW;  D0F748976EBC222C CRC64;

  Query Match             82.6%;  Score 38;  DB 2;  Length 1280;
  Best Local Similarity   75.0%;  Pred. No. 1e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:||| ||
Db        227 HELIHALH 234


RESULT 14
Q8IJV0
ID   Q8IJV0      PRELIMINARY;      PRT;  349 AA.
AC   Q8IJV0;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   PF10_0092.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=22255705; PubMed=12368864;
RA   Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA   Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA   Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA   Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA   Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA   Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA   McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA   Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA   Fraser C.M., Barrell B.;
RT   "Genome sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Nature 419:498-511(2002).
DR   EMBL; AE014830; AAN35290.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   349 AA;  40775 MW;  9D7C20FEACED4464 CRC64;

  Query Match             80.4%;  Score 37;  DB 5;  Length 349;
  Best Local Similarity   50.0%;  Pred. No. 43;
  Matches    4;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |:|:|::|
Db         96 HELVHIVH 103


RESULT 15
Q8X008
ID   Q8X008      PRELIMINARY;      PRT;  707 AA.
AC   Q8X008;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Related to hydroxyproline-rich glycoprotein.
GN   B23H20.050.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA   Nyakatura G., Mewes H.W., Mannhaupt G.;
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   German Neurospora genome project;
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL669988; CAD21077.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
SQ   SEQUENCE   707 AA;  77817 MW;  C49BC3C1A18D83F5 CRC64;

  Query Match             80.4%;  Score 37;  DB 3;  Length 707;
  Best Local Similarity   62.5%;  Pred. No. 87;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HDLIHVLH 8
              |||:::||
Db        158 HDLLYILH 165


Search completed: November 21, 2003, 15:49:45
Job time : 30.5 secs
```

```
Connecting via Winsock to STN



Welcome to STN International!  Enter x:x

LOGINID:ssspta1653hxp

PASSWORD:
TERMINAL (ENTER 1, 2, 3, OR ?):2


* * * * * * * * * *   Welcome to STN International   * * * * * * * * * *

 NEWS  1           Web Page URLs for STN Seminar Schedule - N. America
 NEWS  2           "Ask CAS" for self-help around the clock
 NEWS  3  SEP 09   CA/CAplus records now contain indexing from 1907 to the
                   present
 NEWS  4  AUG 05   New pricing for EUROPATFULL and PCTFULL effective
                   August 1, 2003
 NEWS  5  AUG 13   Field Availability (/FA) field enhanced in BEILSTEIN
 NEWS  6  AUG 18   Data available for download as a PDF in RDISCLOSURE
 NEWS  7  AUG 18   Simultaneous left and right truncation added to PASCAL
 NEWS  8  AUG 18   FROSTI and KOSMET enhanced with Simultaneous Left and Righ
                   Truncation
 NEWS  9  AUG 18   Simultaneous left and right truncation added to ANABSTR
 NEWS 10  SEP 22   DIPPR file reloaded
 NEWS 11  SEP 25   INPADOC: Legal Status data to be reloaded
 NEWS 12  SEP 29   DISSABS now available on STN
 NEWS 13  OCT 10   PCTFULL: Two new display fields added
 NEWS 14  OCT 21   BIOSIS file reloaded and enhanced
 NEWS 15  OCT 28   BIOSIS file segment of TOXCENTER reloaded and enhanced
 NEWS 16  NOV 24   MSDS-CCOHS file reloaded

 NEWS EXPRESS  NOVEMBER 14 CURRENT WINDOWS VERSION IS V6.01c, CURRENT
               MACINTOSH VERSION IS V6.0b(ENG) AND V6.0Jb(JP),
               AND CURRENT DISCOVER FILE IS DATED 23 SEPTEMBER 2003
 NEWS HOURS    STN Operating Hours Plus Help Desk Availability
 NEWS INTER    General Internet Information
 NEWS LOGIN    Welcome Banner and News Items
 NEWS PHONE    Direct Dial and Telecommunication Network Access to STN
 NEWS WWW      CAS World Wide Web Site (general information)

Enter NEWS followed by the item number or name to see news on that
specific topic.

  All use of STN is subject to the provisions of the STN Customer
  agreement.  Please note that this agreement limits use to scientific
  research.  Use for software development or design or implementation
  of commercial gateways or other similar uses is prohibited and may
  result in loss of user privileges and other penalties.

  * * * * * * * * * * * * * * STN  Columbus  * * * * * * * * * * * * * * *

FILE 'HOME' ENTERED AT 13:32:58 ON 25 NOV 2003

=> file medline, uspatful, dgene, embase, wpids, biosis, biobusiness, jicst, fsta'
'FSTA'' IS NOT A VALID FILE NAME
Enter "HELP FILE NAMES" at an arrow prompt (=>) for a list of files
that are available.  If you have requested multiple files, you can
specify a corrected file name or you can enter "IGNORE" to continue
accessing the remaining file names entered.
```

```
ENTER A FILE NAME OR (IGNORE):fsta
COST IN U.S. DOLLARS                                    SINCE FILE      TOTAL
                                                           ENTRY    SESSION
FULL ESTIMATED COST                                         0.42       0.42

FILE 'MEDLINE' ENTERED AT 13:34:15 ON 25 NOV 2003

FILE 'USPATFULL' ENTERED AT 13:34:15 ON 25 NOV 2003
CA INDEXING COPYRIGHT (C) 2003 AMERICAN CHEMICAL SOCIETY (ACS)

FILE 'DGENE' ENTERED AT 13:34:15 ON 25 NOV 2003
COPYRIGHT (C) 2003 THOMSON DERWENT

FILE 'EMBASE' ENTERED AT 13:34:15 ON 25 NOV 2003
COPYRIGHT (C) 2003 Elsevier Inc. All rights reserved.

FILE 'WPIDS' ENTERED AT 13:34:15 ON 25 NOV 2003
COPYRIGHT (C) 2003 THOMSON DERWENT

FILE 'BIOSIS' ENTERED AT 13:34:15 ON 25 NOV 2003
COPYRIGHT (C) 2003 BIOLOGICAL ABSTRACTS INC.(R)

FILE 'BIOBUSINESS' ENTERED AT 13:34:15 ON 25 NOV 2003
COPYRIGHT (C) 2003 Biological Abstracts, Inc. (BIOSIS)

FILE 'JICST-EPLUS' ENTERED AT 13:34:15 ON 25 NOV 2003
COPYRIGHT (C) 2003 Japan Science and Technology Agency (JST)

FILE 'FSTA' ENTERED AT 13:34:15 ON 25 NOV 2003
COPYRIGHT (C) 2003 International Food Information Service

=> s hybrid or fusion protein
L1      507200 HYBRID OR FUSION PROTEIN

=> s IgE and transport of toxin
L2           1 IGE AND TRANSPORT OF TOXIN

=> d l2 ti abs ibib tot

L2   ANSWER 1 OF 1  USPATFULL on STN
TI     Dielectrophoretic separation and immunoassay methods on active
       electronic matrix devices
AB     This invention relates to devices and methods for performing active,
       multi-step molecular and biological sample preparation and diagnostic
       analyses employing immunochemical techniques. It relates generally to
       bioparticle separation, bioparticle enrichment, and electric
       field-mediated immunochemical detection on active electronic matrix
       devices utilizing AC and DC electric fields. More specifically, the
       invention relates to devices and methods for sample
       preparation/manipulation, immunoimmobilization, and immunoassays, all of
       which can be conducted on one or more active electronic chip devices
       within a single system. These manipulations are useful in a variety of
       applications, including, for example, detection of pathogenic bacteria
       and biological warfare agents, point-of-care diagnostics, food or
       medical product quality control assays, and other biological assays.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.
ACCESSION NUMBER:       2003:211475  USPATFULL
TITLE:                  Dielectrophoretic separation and immunoassay methods on
                        active electronic matrix devices
INVENTOR(S):            Huang, Ying, San Diego, CA, UNITED STATES
                        Ewalt, Karla, San Diego, CA, UNITED STATES
                        Haigis, Robert, San Diego, CA, UNITED STATES
                        Forster, Anita H., Santee, CA, UNITED STATES
```

| | |
|---|---|
| | Krihak, Michael K., San Diego, CA, UNITED STATES |
| PATENT ASSIGNEE(S): | NANOGEN, INC. (U.S. corporation) |

| | NUMBER | KIND | DATE | |
|---|---|---|---|---|
| PATENT INFORMATION: | US 2003146100 | A1 | 20030807 | |
| APPLICATION INFO.: | US 2002-72660 | A1 | 20020206 | (10) |

| | |
|---|---|
| DOCUMENT TYPE: | Utility |
| FILE SEGMENT: | APPLICATION |
| LEGAL REPRESENTATIVE: | O'MELVENY & MEYERS, 114 PACIFICA, SUITE 100, IRVINE, CA, 92618 |
| NUMBER OF CLAIMS: | 21 |
| EXEMPLARY CLAIM: | 1 |
| NUMBER OF DRAWINGS: | 14 Drawing Page(s) |
| LINE COUNT: | 1844 |

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

=> s mastocyte and degranulation inhibition
L3 10 MASTOCYTE AND DEGRANULATION INHIBITION

=> d l3 ti abs ibib tot

L3 ANSWER 1 OF 10 USPATFULL on STN
TI Phthalazine derivatives as phosphodiesterase 4 inhibitors
AB Compounds of formula (I) wherein B is alkylene, amino, CONH or a bond; Cy is optionally substituted phenyl or heteroaryl; R is H, phenyl or (C.sub.1-4)alkyl optionally substituted; R.sub.1 is (C.sub.1-6)alkyl or polyfluoro(C.sub.1-6)-alkyl; R.sub.2 is (C.sub.4-7)cycloalkyl optionally containing an oxygen atom and optionally substituted; and the N.fwdarw.O derivatives and pharmaceutically acceptable salt thereof are PDE 4 and TNF.alpha. inhibitors. ##STR1##

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

| | |
|---|---|
| ACCESSION NUMBER: | 2003:184083 USPATFULL |
| TITLE: | Phthalazine derivatives as phosphodiesterase 4 inhibitors |
| INVENTOR(S): | Napoletano, Mauro, Milan, ITALY<br>Norcini, Gabriele, Vizzola Ticino, ITALY<br>Botta, Daniela, Como, ITALY<br>Grancini, Giancarlo, Nova Milanese, ITALY<br>Morazzoni, Gabriele, Lainate, ITALY<br>Santangelo, Francesco, Milan, ITALY<br>Siro Herrero, Jorge G., Cala d'Henares, SPAIN<br>Garcia Navaio, Jose Luis, Madrid, SPAIN<br>Alvarez-Builla, Julio G., Madrid, SPAIN |
| PATENT ASSIGNEE(S): | Zambon Group S.p.A., Vincenza, ITALY (non-U.S. corporation) |

| | NUMBER | KIND | DATE | |
|---|---|---|---|---|
| PATENT INFORMATION: | US 6589951 | B1 | 20030708 | |
| | WO 9932456 | | 19990701 | |
| APPLICATION INFO.: | US 2000-581506 | | 20000810 | (9) |
| | WO 1998-EP8291 | | 19981217 | |

| | NUMBER | DATE |
|---|---|---|
| PRIORITY INFORMATION: | IT 1997-MI2806 | 19971219 |

| | |
|---|---|
| DOCUMENT TYPE: | Utility |
| FILE SEGMENT: | GRANTED |
| PRIMARY EXAMINER: | Ford, John M. |
| LEGAL REPRESENTATIVE: | Arent Fox Kintner Plotkin & Kahn |
| NUMBER OF CLAIMS: | 8 |

EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 0 Drawing Figure(s); 0 Drawing Page(s)
LINE COUNT: 1463
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L3 ANSWER 2 OF 10 USPATFULL on STN
TI Tricyclic phthalazine derivatives as phosphodiesterase 4 inhibitors
AB Tricyclic phthalazine compounds of formula (I) ##STR1##

wherein A is a 5-7 membered heterocycle containing from 1 to 4 nitrogen atoms, optionally partially or totally unsaturated, and optionally substituted by a (C.sub.1-4)alkyl group in turn optionally substituted; Z is NH, methylene, a C.sub.2-6 alkylene chain optionally branched and/or unsaturated and/or interrupted by a C.sub.5-7 cycloalkyl residue; Cy is phenyl or heterocycle optionally substituted by one or more substituents, or a COR.sub.4 group wherein R.sub.4 is hydroxy, alkoxy, amino optionally substituted by one or two (C.sub.1-6)alkyl groups or by hydroxy; R is a (C.sub.1-6)alkyl or polyfluoro(C.sub.1-6)alkyl group; R.sub.1 is hydrogen; a (C.sub.1-8)-alkyl, (C.sub.2-8)-alkenyl or (C.sub.2-8)-alkynyl group optionally substituted by hydroxy, oxo, aryl or heterocycle, and optionally interrupted by one or more heteroatoms or heterogroups; a (C.sub.1-4)alkoxy group or a (C.sub.4-7)cycloalkoxy group optionally containing an oxygen atom and optionally substituted by a polar substituent in the cyclic moiety, aryloxy aryl-(C.sub.1-10)-alkoxy; the N--O derivatives and the pharmaceutically acceptable salts thereof are described. The compounds of formula (I) are PDE 4 inhibitors.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.
ACCESSION NUMBER: 2003:53809 USPATFULL
TITLE: Tricyclic phthalazine derivatives as phosphodiesterase 4 inhibitors
INVENTOR(S): Napoletano, Mauro, Milan, ITALY
Norcini, Gabriele, Vizzola Ticino, ITALY
Pellacini, Franco, Milan, ITALY
Morazzoni, Gabriele, Lainate, ITALY
PATENT ASSIGNEE(S): Zambon Group S.p.A., Vicenza, ITALY (non-U.S. corporation)

| | NUMBER | KIND | DATE | |
|---|---|---|---|---|
| PATENT INFORMATION: | US 6525055 | B1 | 20030225 | |
| | WO 2000026218 | | 20000511 | |
| APPLICATION INFO.: | US 2001-830679 | | 20010430 | (9) |
| | WO 1999-EP7304 | | 19991001 | |

| | NUMBER | DATE |
|---|---|---|
| PRIORITY INFORMATION: | IT 1998-MI2319 | 19981029 |

DOCUMENT TYPE: Utility
FILE SEGMENT: GRANTED
PRIMARY EXAMINER: Raymond, Richard L.
LEGAL REPRESENTATIVE: Oblon, Spivak, McClelland, Maier & Neustadt, P.C.
NUMBER OF CLAIMS: 13
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 0 Drawing Figure(s); 0 Drawing Page(s)
LINE COUNT: 703
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L3 ANSWER 3 OF 10 USPATFULL on STN
TI Phthalazine derivatives phosphodiesterase 4 inhibitors
AB Compounds of formula (I), ##STR1##

wherein {character pullout} is a single or double bond; Z is NH,

methylene, a (C.sub.2-C.sub.6)alkylene chain optionally branched and/or unsaturated and/or interrupted by a (C.sub.5-C.sub.7)cycloalkyl residue; A is phenyl or heterocycle optionally substituted or a COR.sub.4 group wherein R.sub.4 is hydroxy, (C.sub.1-C.sub.6)-alkoxy, amino optionally substituted; R is a (C.sub.1-C.sub.6)alkyl or polyfluoro(C.sub.1-C.sub.6)alkyl group; R.sub.1 is absent when {character pullout} is a double bond or, when {character pullout} is a single bond, is (a) hydrogen; (b) (C.sub.1-C.sub.6)alkyl optionally substituted; (c) --COR.sub.6 wherein R.sub.6 is hydrogen, aryl, aryl-(C.sub.1-C.sub.6)alkyl, amino optionally alkylated or monohydroxylated, hydroxy, (C.sub.1-C.sub.4)alkoxy, carboxy, (C.sub.1-C.sub.4)alkoxycarbonyl, formula (1), or (C.sub.1-C.sub.4)alkyl optionally substituted by heterocycle; (d) (C.sub.1-C.sub.4)-alkylsulfonyl; R.sub.2 represents two hydrogen atoms or a group --O when {character pullout} is a single bond, or, when {character pullout} is a double bond, R.sub.2 is hydrogen, cyano, (C.sub.1-C.sub.4)alkoxycarbonyl, amido, optionally substituted aryl or heterocycle, (C.sub.1-C.sub.8)alkyl, (C.sub.2-C.sub.8)alkenyl or (C.sub.2-C.sub.8)alkynyl optionally branched and/or substituted; aryloxy, heterocyclyloxy, aryl-(C.sub.1-C.sub.4)alkoxy, heterocyclyloxy-(C.sub.1-C.sub.4)alkoxy, amino substituted by one or two (C.sub.1-C.sub.4)-alkyl group(s), arylamino, heterocyclylamino, aryl-(C.sub.1-C.sub.4)alkylamino, heterocyclyl-(C.sub.1-C.sub.4)-alkylamino; R.sub.3 is hydrogen, or a (C.sub.1-C.sub.8)alkyl, (C.sub.2-C.sub.8)alkenyl or (C.sub.2-C.sub.8)alkynyl group optionally substituted, and optionally interrupted; the N.fwdarw.O derivatives of the compounds of formula (I) and the pharmaceutically acceptable salts thereof are PDE 4 inhibitors.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

| | |
|---|---|
| ACCESSION NUMBER: | 2002:326004 USPATFULL |
| TITLE: | Phthalazine derivatives phosphodiesterase 4 inhibitors |
| INVENTOR(S): | Napoletano, Mauro, Milan, ITALY<br>Norcini, Gabriele, Vizzola Ticino, ITALY<br>Grancini, Giancarlo, Nova Milanese, ITALY<br>Pellacini, Franco, Milan, ITALY<br>Leali, Gian Marco, Milan, ITALY<br>Morazzoni, Gabriele, Lainate, ITALY |
| PATENT ASSIGNEE(S): | Zambon Group S.p.A., Vicenza, ITALY (non-U.S. corporation) |

| | NUMBER | KIND | DATE |
|---|---|---|---|
| PATENT INFORMATION: | US 6492360 | B1 | 20021210 |
| APPLICATION INFO.: | US 2001-976436 | | 20011015 (9) |
| RELATED APPLN. INFO.: | Continuation of Ser. No. US 743813, now patented, Pat. No. US 6329370 | | |

| | NUMBER | DATE |
|---|---|---|
| PRIORITY INFORMATION: | IT 1998-MI1670 | 19980721 |

| | |
|---|---|
| DOCUMENT TYPE: | Utility |
| FILE SEGMENT: | GRANTED |
| PRIMARY EXAMINER: | Raymond, Richard L. |
| LEGAL REPRESENTATIVE: | Oblon, Spivak, McClelland, Maier & Neustadt, P.C. |
| NUMBER OF CLAIMS: | 17 |
| EXEMPLARY CLAIM: | 1 |
| NUMBER OF DRAWINGS: | 0 Drawing Figure(s); 0 Drawing Page(s) |
| LINE COUNT: | 3022 |

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L3 ANSWER 4 OF 10 USPATFULL on STN

TI Phthalazine derivatives as phosphodiesterase 4 inhibitors

AB The present invention relates to phthalazine derivatives, pharmaceutical compositions containing them, and to their use as phosphodiesterase 4

inhbitors.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

| | |
|---|---|
| ACCESSION NUMBER: | 2002:112921 USPATFULL |
| TITLE: | Phthalazine derivatives as phosphodiesterase 4 inhibitors |
| INVENTOR(S): | Napoletano, Mauro, Milan, ITALY<br>Norcini, Gabriele, Varese, ITALY<br>Grancini, Giancarlo, Milan, ITALY<br>Pellacini, Franco, Milan, ITALY<br>Morazzoni, Gabriele, Milan, ITALY |
| PATENT ASSIGNEE(S): | ZAMBON GROUP S.P.A., Vicenze, ITALY, 36100 (non-U.S. corporation) |

| | NUMBER | KIND | DATE |
|---|---|---|---|
| PATENT INFORMATION: | US 2002058662 | A1 | 20020516 |
| | US 6498160 | B2 | 20021224 |
| APPLICATION INFO.: | US 2001-987266 | A1 | 20011114 (9) |
| RELATED APPLN. INFO.: | Continuation of Ser. No. US 2001-764983, filed on 22 Jan 2001, PATENTED | | |

| | NUMBER | DATE |
|---|---|---|
| PRIORITY INFORMATION: | IT 1998-MI1671 | 19980721 |
| | WO 1999-EP5068 | 19990716 |

| | |
|---|---|
| DOCUMENT TYPE: | Utility |
| FILE SEGMENT: | APPLICATION |
| LEGAL REPRESENTATIVE: | OBLON SPIVAK MCCLELLAND MAIER & NEUSTADT PC, FOURTH FLOOR, 1755 JEFFERSON DAVIS HIGHWAY, ARLINGTON, VA, 22202 |
| NUMBER OF CLAIMS: | 6 |
| EXEMPLARY CLAIM: | 1 |
| LINE COUNT: | 1105 |

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L3 ANSWER 5 OF 10 USPATFULL on STN
TI Benzazine derivatives as phosphodiesterase 4 inhibitors
AB Compounds of formula I: ##STR1##

wherein A is a heterocycle containing a nitrogen atom and optionally saturated or unsaturated and optionally further substituted by an oxo group (.dbd.O); R is: hydrogen, cyano, (C.sub.1-4)alkoxycarbonyl, carbamoyl; optionally substituted (C.sub.4-7)-cycloalkyl, aryl or heterocycle; (C.sub.1-8)alkyl, (C.sub.2-8)alkenyl or (C.sub.2-8)alkynyl optionally branched and/or substituted by (C.sub.4-7) cycloalkyl, aryl or heterocycle; aryloxy, heterocyclyloxy, aryl(C.sub.1-4)alkoxy, heterocyclyl(C.sub.1-4)alkoxy, amino substituted by one or two (C.sub.1-4)alkyl group(s), aryl-amino, heterocyclyl-amino, aryl(C.sub.1-4)alkyl-amino, or heterocyclyl(C.sub.1-4)alkylamino; Y is methylene or ethylene; W is an optionally substituted aryl or heterocycle; R.sub.1 is hydrogen, (C.sub.4-7)cycloalkyl or a (C.sub.2-8)alkyl, (C.sub.2-8)alkenyl or (C.sub.2-8)alkynyl group optionally substituted by hydroxy, oxo, (C.sub.4-7)cycloalkyl, aryl or heterocycle, and optionally interrupted by one or more heteroatom(s) or heterogroup(s); R.sub.2 is a (C.sub.1-6)alkyl or polyfluoro(C.sub.1-6)alkyl group; the N.fwdarw.O derivatives of the compounds of formula I and the pharmaceutically acceptable salts thereof. The compounds of formula (I) are PDE 4 inhibitors and may be used in compositions and methods involving PDE 4 inhibition.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

| | |
|---|---|
| ACCESSION NUMBER: | 2002:57805 USPATFULL |
| TITLE: | Benzazine derivatives as phosphodiesterase 4 inhibitors |

INVENTOR(S): Napoletano, Mauro, Milan, ITALY
Norcini, Gabriele, Vizzola Ticino, ITALY
Grancini, Giancarlo, Nova Milanese, ITALY
Pellacini, Franco, Milan, ITALY
Morazzoni, Gabriele, Lainate, ITALY
Pradella, Lorenzo, Cernusco sul Naviglio, ITALY
PATENT ASSIGNEE(S): Zambon Group S.p.A., Vicenza, ITALY (non-U.S. corporation)

| | NUMBER | KIND | DATE | |
|---|---|---|---|---|
| PATENT INFORMATION: | US 6358973 | B1 | 20020319 | |
| | WO 2000021947 | | 20000420 | |
| APPLICATION INFO.: | US 2001-806496 | | 20010413 | (9) |
| | WO 1999-EP7302 | | 19991010 | |
| | | | 20010413 | PCT 371 date |

| | NUMBER | DATE |
|---|---|---|
| PRIORITY INFORMATION: | IT 1998-MI2216 | 19981015 |

DOCUMENT TYPE: Utility
FILE SEGMENT: GRANTED
PRIMARY EXAMINER: Davis, Zinna Northington
LEGAL REPRESENTATIVE: Oblon, Spivak, McClelland, Maier & Neustadt, P.C.
NUMBER OF CLAIMS: 19
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 0 Drawing Figure(s); 0 Drawing Page(s)
LINE COUNT: 1184
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L3 ANSWER 6 OF 10 USPATFULL on STN
TI Phthalazine derivatives as phosphodiesterase 4 inhibitors
AB The present invention provides a compound selected from the group including: N-3-acetyl-1-(3,5-dichloropyridin-4-ylmethyl)-5-cyclopentyloxy-6-methoxy-4H-phthalazine; 6,7-dimethoxy-1-pyridin-4-ylmethyl-4-thiazol-2-yl-phthalazine; 1-(6,7-dimethoxy-4-pyridin-4-ylmethyl-1H-phthalazin-2-yl)ethanone; 2-methanesulphonyl-6,7-dimethoxy-4-pyridin-4-ylmethyl-1,2-dihydrophthalazine; 2-formyl-6,7-dimethoxy-4-pyridin-4-ylmethyl-1,2-dihydrophthalazine; 1-(6,7-dimethoxy-4-pyridin-4-ylmethyl-1H-phthalazin-2-yl)-1-imidazol-1-ylmethanone; 1-(3,5-dichloro-pyridin-4-ylmethyl)-3-methansulphonyl-6-difluoromethoxy-5-(tetrahydro-furan-2-yloxy)-4H-phthalazine; N.fwdarw.O derivatives thereof; and pharmaceutically acceptable salts thereof. The invention also provides a pharmaceutical composition, which contains a therapeutically effective amount of the above compound in admixture with a pharmaceutically acceptable carrier.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.
ACCESSION NUMBER: 2002:14001 USPATFULL
TITLE: Phthalazine derivatives as phosphodiesterase 4 inhibitors
INVENTOR(S): Napoletano, Mauro, Milan, ITALY
Norcini, Gabriele, Vizzola Ticino, ITALY
Grancini, Giancarlo, Nova Milanese, ITALY
Pellacini, Franco, Milan, ITALY
Morazzoni, Gabriele, Lainate, ITALY
PATENT ASSIGNEE(S): Zambon Group S.p.A., Vicenza, ITALY (non-U.S. corporation)

| | NUMBER | KIND | DATE | |
|---|---|---|---|---|
| PATENT INFORMATION: | US 6340684 | B1 | 20020122 | |
| | WO 2000005219 | | 20000203 | |
| APPLICATION INFO.: | US 2001-764983 | | 20010122 | (9) |

| | | |
|---|---|---|
| | WO 1999-EP5068 | 19990716 |
| | | 20010122 PCT 371 date |

| | NUMBER | DATE |
|---|---|---|
| PRIORITY INFORMATION: | IT 1998-MI1671 | 19980721 |

DOCUMENT TYPE: Utility
FILE SEGMENT: GRANTED
PRIMARY EXAMINER: Bernhardt, Emily
LEGAL REPRESENTATIVE: Oblon, Spivak, McClelland, Maier & Neustadt, P.C.
NUMBER OF CLAIMS: 9
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 0 Drawing Figure(s); 0 Drawing Page(s)
LINE COUNT: 1076
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L3 ANSWER 7 OF 10 USPATFULL on STN
TI Phthalazine derivatives phosphodiesterase 4 inhibitors
AB The present invention provides a compound selected from the group including: 1-(3,5-dichloro-pyridin-4-ylmethyl)-6-methoxy-4-phenyl-phthalazine; 4-(3,5-dichloro-pyridin-4-ylmethyl)-7-methoxy-1H-phthalazin-2-carboxylic acid methyl ester; benzyl-{3-{1-(3,5-dichloro-pyridin-4-ylmethyl)-6-methoxy-phthalazin-5-yl}-prop-2-ynyl}-methyl-amine; 1-(3,5-dichloro-pyridin-4-ylmethyl)-6-methoxy-5-(5-morpholin-4-yl-pent-1-ynyl)-phthalazine dihydrochloride; 3-{1-(3,5-dichloro-pyridin-4-ylmethyl)-6-methoxy-phthalazin-5-yl}-prop-2-yn-1-ol; 1-(3,5-dichloro-pyridin-4-ylmethyl)-6-methoxy-4-morpholin-4-yl-phthalazine; 1-(3,5-dichloro-pyridin-4-ylmethyl)-6-methoxy-4-(1,2,4)triazol-1-yl-phthalazine; N.fwdarw.O derivatives thereof; and pharmaceutically acceptable salts thereof. The invention also provides a pharmaceutical composition which includes a therapeutically effective amount of the above compound in admixture with a suitable carrier.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.
ACCESSION NUMBER: 2001:226627 USPATFULL
TITLE: Phthalazine derivatives phosphodiesterase 4 inhibitors
INVENTOR(S): Napoletano, Mauro, Milan, Italy
Norcini, Gabriele, Vizzola Ticino, Italy
Grancini, Giancarlo, Nova Milanese, Italy
Pellacini, Franco, Milan, Italy
Leali, Gian Marco, Milan, Italy
Morazzoni, Gabriele, Lainate, Italy
PATENT ASSIGNEE(S): Zambon Group S.p.A., Vicenza, Italy (non-U.S. corporation)

| | NUMBER | KIND | DATE | |
|---|---|---|---|---|
| PATENT INFORMATION: | US 6329370 | B1 | 20011211 | |
| | WO 2000005218 | | 20000203 | |
| APPLICATION INFO.: | US 2001-743813 | | 20010122 | (9) |
| | WO 1999-EP4904 | | 19990713 | |
| | | | 20010122 | PCT 371 date |
| | | | 20010122 | PCT 102(e) date |

| | NUMBER | DATE |
|---|---|---|
| PRIORITY INFORMATION: | IT 1998-MI1670 | 19980721 |

DOCUMENT TYPE: Utility
FILE SEGMENT: GRANTED
PRIMARY EXAMINER: Bernhardt, Emily
LEGAL REPRESENTATIVE: Oblon, Spivak, McClelland, Maier & Neustadt, P.C.
NUMBER OF CLAIMS: 2
EXEMPLARY CLAIM: 1
LINE COUNT: 2920

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L3 ANSWER 8 OF 10 USPATFULL on STN
TI Benzazine derivatives phosphodiesterase 4 inhibitors
AB ##STR1##

Compounds of formula (I) wherein A is an orthocondensed heterocycle optionally substituted by certain substituents and necessarily substituted by a --B--Cy group where the variables are as defined in the specification and the N.fwdarw.O derivatives and pharmaceutically acceptable salts thereof are phosphodiesterase-4 inhibitors.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

| | | | | |
|---|---|---|---|---|
| ACCESSION NUMBER: | 2001:168141 USPATFULL | | | |
| TITLE: | Benzazine derivatives phosphodiesterase 4 inhibitors | | | |
| INVENTOR(S): | Napoletano, Mauro, Milan, Italy<br>Norcini, Gabriele, Varese, Italy<br>Botta, Daniela, Como, Italy<br>Grancini, Giancarlo, Milan, Italy<br>Morazzoni, Gabriele, Milan, Italy | | | |
| PATENT ASSIGNEE(S): | Zambon Group S.p.A., Vicenza, Italy (non-U.S. corporation) | | | |
| | NUMBER | KIND | DATE | |
| PATENT INFORMATION: | US 6297257 | B1 | 20011002 | |
| | WO 9932449 | | 19990701 | |
| APPLICATION INFO.: | US 2000-581505 | | 20000713 | (9) |
| | WO 1998-EP8292 | | 19981217 | |
| | | | 20000713 | PCT 371 date |
| | | | 20000713 | PCT 102(e) date |
| | NUMBER | | DATE | |
| PRIORITY INFORMATION: | IT 1997-MI2807 | | 19971219 | |
| DOCUMENT TYPE: | Utility | | | |
| FILE SEGMENT: | GRANTED | | | |
| PRIMARY EXAMINER: | Dentz, Bernard | | | |
| LEGAL REPRESENTATIVE: | Arent Fox Kintner Plotkin & Kahn PLLC | | | |
| NUMBER OF CLAIMS: | 6 | | | |
| EXEMPLARY CLAIM: | 1 | | | |
| LINE COUNT: | 1248 | | | |

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L3 ANSWER 9 OF 10 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
TI Therapeutic mechanism of a Chinese medicine decoction for urticaria.
AB Objective To investigate therapeutic mechanism of a Chinese medicine decoction for urticaria. Methods **Mastocyte** degranulation test and passive skin allergy test were conducted in animal model. A sandwich ELISA technique was applied to detect serum IL - 2 and IL - 4 in patients with acute urticaria before and after oral administration of the decoction. Results Significant inhibition of **mastocyte** degranulation was found in mice taken the decoction in comparison with controls. There were significantly increased level of IL - 2 and reduced level of IL - 4 in sera of patients with acute urticaria. Serum levels of IL - 2 and IL - 4 recovered to normal in patients after taking the decoction. Conclusion The Chinese medicine decoction appears to stabilize the membrane of mastocytes and inhibit antigen - antibody binding. The decoction is also likely to adjust serum levels of IL - 2 and IL - 4 to normal.

| | |
|---|---|
| ACCESSION NUMBER: | 2000:500146 BIOSIS |
| DOCUMENT NUMBER: | PREV200000500267 |
| TITLE: | Therapeutic mechanism of a Chinese medicine decoction for urticaria. |

AUTHOR(S): Song Dongyan [Reprint author]; Chen Deyu [Reprint author]; Xiong Xia [Reprint author]
CORPORATE SOURCE: Department of Dermatology, Affiliated Hospital, Zhenjiang Medical College, Jiangsu, 212001, China
SOURCE: Zhonghua Pifuke Zazhi, (August, 2000) Vol. 33, No. 4, pp. 251-253. print.
CODEN: CHFTAJ. ISSN: 0412-4030.
DOCUMENT TYPE: Article
LANGUAGE: Chinese
ENTRY DATE: Entered STN: 15 Nov 2000
Last Updated on STN: 11 Jan 2002

L3 ANSWER 10 OF 10 JICST-EPlus COPYRIGHT 2003 JST on STN
TI Regulation of inflammation in urticaria.
AB Urticaria is characterized by wheal and flare, which usually disappear within several hours. However, there are some cases with severe and/or prolonged eruptions lasting more than 24 hours and resistant to treatments by ordinary histamine antagonists. Histology of such cases accompanies various cell infiltrations, such as eosinophils and neutrophils. In order to know the mechanism of such infiltrations, we studied the release of chemotactic activities from skin slices for neutrophils and eosinophils. The factor was identified with LTB4 by HPLC analysis and released either by antigens or substance P. Moreoever, studies with skin chambers on patients with chronic urticaria revealed spontaneous release of substance P, suggesting the involvement of neuropeptides in the pathogenesis of chronic urticaria. Incubation of skin slices with dexamethasone scarcely inhibited histamine release, but almost abolished release of LTB4 in response to antigens. Recently-developed histamine H1-antagonists, called anti-allergic drugs, inhibited both degranulation and TNF.ALPHA. releases from rat mast cell (RBL-2H3) line, but concentrations required to inhibit TNF.ALPHA. release was about one tenth of those to inhibit **degranulation**. **Inhibition** of LTB4 and TNF.ALPHA. from mast cells may partially account for clinical efficacy of corticosteroids and some anti-allergic drugs for treatment of chronic idiopathic urticaria. (author abst.)
ACCESSION NUMBER: 980978358 JICST-EPlus
TITLE: Regulation of inflammation in urticaria.
AUTHOR: HIDE MICHIHIRO; TANAKA TOSHIHIKO; KORO OSAMU; YAMAMOTO SHOSO
CORPORATE SOURCE: Hiroshima Univ., Sch. of Med.
SOURCE: Ensho (Japanese Journal of Inflammation), (1998) vol. 18, no. 5, pp. 349-354. Journal Code: Y0899A (Fig. 2, Ref. 26)
CODEN: ENSHEE; ISSN: 0389-4290
PUB. COUNTRY: Japan
DOCUMENT TYPE: Journal; Commentary
LANGUAGE: Japanese
STATUS: New

=> d his

(FILE 'HOME' ENTERED AT 13:32:58 ON 25 NOV 2003)

FILE 'MEDLINE, USPATFULL, DGENE, EMBASE, WPIDS, BIOSIS, BIOBUSINESS, JICST-EPLUS, FSTA' ENTERED AT 13:34:15 ON 25 NOV 2003
L1 507200 S HYBRID OR FUSION PROTEIN
L2 1 S IGE AND TRANSPORT OF TOXIN
L3 10 S MASTOCYTE AND DEGRANULATION INHIBITION

=> s clostridium botulinum toxin
L4 514 CLOSTRIDIUM BOTULINUM TOXIN

=> s l4 and allergic response
L5 0 L4 AND ALLERGIC RESPONSE

```
=> s toxin and allergy
L6        3536 TOXIN AND ALLERGY

=> s l6 and allergic response
L7         229 L6 AND ALLERGIC RESPONSE

=> s l7 and tetanus toxin
L8          27 L7 AND TETANUS TOXIN

=> d l8 ti abs ibib tot

L8   ANSWER 1 OF 27  USPATFULL on STN
TI     Bi-directionally cloned random cDNA expression vector libraries,
       compositions and methods of use
AB     The present invention provides random cDNA expression vector libraries,
       comprising expression vectors which comprise random cDNAs positioned in
       sense and antisense orientation, which are useful for the delivery and
       expression of a combination of genetic effector types to host cells.
       Methods for producing these libraries through bi-directional cloning of
       random cDNAs are also provided. Also provided herein are methods of
       using these libraries to screen for agents capable of modulating cell
       phenotype in desirable ways.

ACCESSION NUMBER:         2003:300312  USPATFULL
TITLE:                    Bi-directionally cloned random cDNA expression vector
                          libraries, compositions and methods of use
INVENTOR(S):              Lorens, James, Portola Valley, CA, UNITED STATES
                          Bogenberger, Jakob M., San Francisco, CA, UNITED STATES

                               NUMBER        KIND    DATE
                          --------------- ------ --------
PATENT INFORMATION:       US 2003211535      A1   20031113
APPLICATION INFO.:        US 2002-142648     A1   20020508  (10)
DOCUMENT TYPE:            Utility
FILE SEGMENT:             APPLICATION
LEGAL REPRESENTATIVE:     BOZICEVIC, FIELD & FRANCIS LLP, 200 MIDDLEFIELD RD,
                          SUITE 200, MENLO PARK, CA, 94025
NUMBER OF CLAIMS:         20
EXEMPLARY CLAIM:          1
NUMBER OF DRAWINGS:       5 Drawing Page(s)
LINE COUNT:               3910

L8   ANSWER 2 OF 27  USPATFULL on STN
TI     Directionally cloned random cDNA expression vector libraries,
       compositions and methods of use
AB     The present invention provides random cDNA expression vector libraries,
       comprising expression vectors which comprise random cDNAs positioned in
       sense orientation. Also provided are random cDNA expression vector
       libraries, comprising expression vectors which comprise random cDNAs
       positioned in antisense orientation. Methods for producing these
       libraries through directional cloning of random cDNAs are also provided.
       Also provided herein are methods of using these libraries to screen for
       agents capable of modulating cell phenotype in desirable ways.

ACCESSION NUMBER:         2003:300239  USPATFULL
TITLE:                    Directionally cloned random cDNA expression vector
                          libraries, compositions and methods of use
INVENTOR(S):              Shen, Mary, Newark, CA, UNITED STATES
                          Yu, Simon, Newark, CA, UNITED STATES
                          Wu, Xian, Redwood City, CA, UNITED STATES
                          Payan, Donald, Hillsborough, CA, UNITED STATES

                               NUMBER        KIND    DATE
```

| | NUMBER | KIND | DATE | |
|---|---|---|---|---|
| PATENT INFORMATION: | US 2003211462 | A1 | 20031113 | |
| APPLICATION INFO.: | US 2002-142662 | A1 | 20020508 | (10) |

DOCUMENT TYPE: Utility
FILE SEGMENT: APPLICATION
LEGAL REPRESENTATIVE: BOZICEVIC, FIELD & FRANCIS LLP, 200 MIDDLEFIELD RD, SUITE 200, MENLO PARK, CA, 94025
NUMBER OF CLAIMS: 26
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 4 Drawing Page(s)
LINE COUNT: 3873

L8 ANSWER 3 OF 27 USPATFULL on STN
TI Human cDNAs and proteins and uses thereof
AB The invention concerns GENSET polynucleotides and polypeptides. Such GENSET products may be used as reagents in forensic analyses, as chromosome markers, as tissue/cell/organelle-specific markers, in the production of expression vectors. In addition, they may be used in screening and diagnosis assays for abnormal GENSET expression and/or biological activity and for screening compounds that may be used in the treatment of GENSET-related disorders.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.
ACCESSION NUMBER: 2003:282611 USPATFULL
TITLE: Human cDNAs and proteins and uses thereof
INVENTOR(S): Bejanin, Stephane, Paris, FRANCE
Tanaka, Hiroaki, Antony, FRANCE
PATENT ASSIGNEE(S): GENSET, S.A., Paris, FRANCE (non-U.S. corporation)

| | NUMBER | KIND | DATE | |
|---|---|---|---|---|
| PATENT INFORMATION: | US 2003198954 | A1 | 20031023 | |
| APPLICATION INFO.: | US 2001-1142 | A1 | 20011114 | (10) |

RELATED APPLN. INFO.: Division of Ser. No. US 2001-924340, filed on 6 Aug 2001, PENDING

| PRIORITY INFORMATION: | NUMBER | DATE | |
|---|---|---|---|
| | WO 2001-IB1715 | 20010806 | |
| | US 2001-305456P | 20010713 | (60) |
| | US 2001-302277P | 20010629 | (60) |
| | US 2001-298698P | 20010615 | (60) |
| | US 2001-293574P | 20010525 | (60) |

DOCUMENT TYPE: Utility
FILE SEGMENT: APPLICATION
LEGAL REPRESENTATIVE: SALIWANCHIK LLOYD & SALIWANCHIK, A PROFESSIONAL ASSOCIATION, 2421 N.W. 41ST STREET, SUITE A-1, GAINESVILLE, FL, 326066669
NUMBER OF CLAIMS: 13
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 4 Drawing Page(s)
LINE COUNT: 25681
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L8 ANSWER 4 OF 27 USPATFULL on STN
TI Human cDNAs and proteins and uses thereof
AB The invention concerns GENSET polynucleotides and polypeptides. Such GENSET products may be used as reagents in forensic analyses, as chromosome markers, as tissue/cell/organelle-specific markers, in the production of expression vectors. In addition, they may be used in screening and diagnosis assays for abnormal GENSET expression and/or biological activity and for screening compounds that may be used in the treatment of GENSET-related disorders.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

| | |
|---|---|
| ACCESSION NUMBER: | 2003:244219 USPATFULL |
| TITLE: | Human cDNAs and proteins and uses thereof |
| INVENTOR(S): | Bejanin, Stephane, Paris, FRANCE<br>Tanaka, Hiroaki, Antony, FRANCE |
| PATENT ASSIGNEE(S): | GENSET, S.A., Paris, FRANCE (non-U.S. corporation) |

| | NUMBER | KIND | DATE |
|---|---|---|---|
| PATENT INFORMATION: | US 2003170628 | A1 | 20030911 |
| APPLICATION INFO.: | US 2001-999570 | A1 | 20011114 (9) |

RELATED APPLN. INFO.: Division of Ser. No. US 2001-924340, filed on 6 Aug 2001, PENDING

| | NUMBER | DATE |
|---|---|---|
| PRIORITY INFORMATION: | WO 2001-IB1715 | 20010806 |
| | US 2001-305456P | 20010713 (60) |
| | US 2001-302277P | 20010629 (60) |
| | US 2001-298698P | 20010615 (60) |
| | US 2001-293574P | 20010525 (60) |

| | |
|---|---|
| DOCUMENT TYPE: | Utility |
| FILE SEGMENT: | APPLICATION |
| LEGAL REPRESENTATIVE: | SALIWANCHIK LLOYD & SALIWANCHIK, A PROFESSIONAL ASSOCIATION, 2421 N.W. 41ST STREET, SUITE A-1, GAINESVILLE, FL, 326066669 |
| NUMBER OF CLAIMS: | 13 |
| EXEMPLARY CLAIM: | 1 |
| NUMBER OF DRAWINGS: | 4 Drawing Page(s) |
| LINE COUNT: | 25549 |

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L8 ANSWER 5 OF 27 USPATFULL on STN

TI Human cDNAs and proteins and uses thereof

AB The invention concerns GENSET polynucleotides and polypeptides. Such GENSET products may be used as reagents in forensic analyses, as chromosome markers, as tissue/cell/organelle-specific markers, in the production of expression vectors. In addition, they may be used in screening and diagnosis assays for abnormal GENSET expression and/or biological activity and for screening compounds that may be used in the treatment of GENSET-related disorders.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

| | |
|---|---|
| ACCESSION NUMBER: | 2003:231986 USPATFULL |
| TITLE: | Human cDNAs and proteins and uses thereof |
| INVENTOR(S): | Bejanin, Stephane, Paris, FRANCE<br>Tanaka, Hiroaki, Antony, FRANCE |
| PATENT ASSIGNEE(S): | GENSET, S.A., Paris, FRANCE (non-U.S. corporation) |

| | NUMBER | KIND | DATE |
|---|---|---|---|
| PATENT INFORMATION: | US 2003162186 | A1 | 20030828 |
| APPLICATION INFO.: | US 2002-154678 | A1 | 20020522 (10) |

| | NUMBER | DATE |
|---|---|---|
| PRIORITY INFORMATION: | US 2001-293574P | 20010525 (60) |
| | US 2001-298698P | 20010615 (60) |
| | US 2001-302277P | 20010629 (60) |
| | US 2001-305456P | 20010713 (60) |

| | |
|---|---|
| DOCUMENT TYPE: | Utility |
| FILE SEGMENT: | APPLICATION |
| LEGAL REPRESENTATIVE: | SALIWANCHIK LLOYD & SALIWANCHIK, A PROFESSIONAL ASSOCIATION, 2421 N.W. 41ST STREET, SUITE A-1, |

GAINESVILLE, FL, 326066669
NUMBER OF CLAIMS: 13
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 4 Drawing Page(s)
LINE COUNT: 25533
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L8 ANSWER 6 OF 27 USPATFULL on STN
TI Human cDNAs and proteins and uses thereof
AB The invention concerns GENSET polynucleotides and polypeptides. Such GENSET products may be used as reagents in forensic analyses, as chromosome markers, as tissue/cell/organelle-specific markers, in the production of expression vectors. In addition, they may be used in screening and diagnosis assays for abnormal GENSET expression and/or biological activity and for screening compounds that may be used in the treatment of GENSET-related disorders.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.
ACCESSION NUMBER: 2003:225673 USPATFULL
TITLE: Human cDNAs and proteins and uses thereof
INVENTOR(S): Bejanin, Stephane, Paris, FRANCE
Tanaka, Hiroaki, Antony, FRANCE
PATENT ASSIGNEE(S): GENSET, S.A., Paris, FRANCE (non-U.S. corporation)

| | NUMBER | KIND | DATE |
|---|---|---|---|
| PATENT INFORMATION: | US 2003157485 | A1 | 20030821 |
| APPLICATION INFO.: | US 2001-992095 | A1 | 20011113 (9) |

RELATED APPLN. INFO.: Division of Ser. No. US 2001-924340, filed on 6 Aug 2001, PENDING

| | NUMBER | DATE |
|---|---|---|
| PRIORITY INFORMATION: | WO 2001-IB1715 | 20010806 |
| | US 2001-305456P | 20010713 (60) |
| | US 2001-302277P | 20010629 (60) |
| | US 2001-298698P | 20010615 (60) |
| | US 2001-293574P | 20010525 (60) |

DOCUMENT TYPE: Utility
FILE SEGMENT: APPLICATION
LEGAL REPRESENTATIVE: SALIWANCHIK LLOYD & SALIWANCHIK, A PROFESSIONAL ASSOCIATION, 2421 N.W. 41ST STREET, SUITE A-1, GAINESVILLE, FL, 326066669
NUMBER OF CLAIMS: 13
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 4 Drawing Page(s)
LINE COUNT: 25484
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L8 ANSWER 7 OF 27 USPATFULL on STN
TI Direct selection of antigen-specific T cells, compositions obtained thereby and methods of use thereof
AB The invention provides a method for convenient analysis and cell separation of antigen-specific T cells based on one or more products secreted by these cells in response to antigen stimulation. The T cells are provided with a capture moiety for the product, which can then be used directly as a label in some instances, or the bound product can be further labeled via label moieties that bind specifically to the product and that are labeled with traditional labeling materials such as fluorophores, radioactive isotopes, chromophores or magnetic particles. The labeled cells are then separated using standard cell sorting techniques based on these labels. Such techniques include flow cytometry, magnetic gradient separation, centrifugation, and the like.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

| | |
|---|---|
| ACCESSION NUMBER: | 2003:155560 USPATFULL |
| TITLE: | Direct selection of antigen-specific T cells, compositions obtained thereby and methods of use thereof |
| INVENTOR(S): | Assenmacher, Mario, Bergisch Gladbach, GERMANY, FEDERAL REPUBLIC OF<br>Miltenyi, Stefan, Bergisch Gladbach, GERMANY, FEDERAL REPUBLIC OF<br>Schmitz, Jurgen, Bergisch Gladbach, GERMANY, FEDERAL REPUBLIC OF |
| PATENT ASSIGNEE(S): | Miltenyi Biotech GmbH, GERMANY, FEDERAL REPUBLIC OF (non-U.S. corporation) |

| | NUMBER | KIND | DATE | |
|---|---|---|---|---|
| PATENT INFORMATION: | US 6576428 | B1 | 20030610 | |
| APPLICATION INFO.: | US 1999-309199 | | 19990510 | (9) |

| | NUMBER | DATE |
|---|---|---|
| PRIORITY INFORMATION: | US 1998-85136P | 19980511 (60) |

| | |
|---|---|
| DOCUMENT TYPE: | Utility |
| FILE SEGMENT: | GRANTED |
| PRIMARY EXAMINER: | Chan, Christina |
| ASSISTANT EXAMINER: | Belyavskyi, Michail |
| LEGAL REPRESENTATIVE: | Morrison & Foerster LLP |
| NUMBER OF CLAIMS: | 19 |
| EXEMPLARY CLAIM: | 1 |
| NUMBER OF DRAWINGS: | 55 Drawing Figure(s); 15 Drawing Page(s) |
| LINE COUNT: | 2084 |

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L8 ANSWER 8 OF 27 USPATFULL on STN

TI Retroviral vectors with separation sequences

AB The invention relates to retroviral vectors comprising fusion nucleic acids useful for expressing a plurality of separate proteins products encoded by genes of interest. The invention further relates to use of the compositions in methods for screening for candidate agents producing an altered phenotype in cells.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

| | |
|---|---|
| ACCESSION NUMBER: | 2003:146187 USPATFULL |
| TITLE: | Retroviral vectors with separation sequences |
| INVENTOR(S): | Lorens, James B., Portola Valley, CA, UNITED STATES<br>Ferrick, David A., El Macero, CA, UNITED STATES |

| | NUMBER | KIND | DATE | |
|---|---|---|---|---|
| PATENT INFORMATION: | US 2003099932 | A1 | 20030529 | |
| APPLICATION INFO.: | US 2002-139146 | A1 | 20020503 | (10) |

| | |
|---|---|
| RELATED APPLN. INFO.: | Continuation-in-part of Ser. No. US 2001-966976, filed on 27 Sep 2001, PENDING Continuation of Ser. No. US 2001-963206, filed on 25 Sep 2001, PENDING Continuation of Ser. No. US 2001-963247, filed on 25 Sep 2001, PENDING Division of Ser. No. US 1998-76624, filed on 12 May 1998, PENDING |
| DOCUMENT TYPE: | Utility |
| FILE SEGMENT: | APPLICATION |
| LEGAL REPRESENTATIVE: | DORSEY & WHITNEY LLP, INTELLECTUAL PROPERTY DEPARTMENT, 4 EMBARCADERO CENTER, SUITE 3400, SAN FRANCISCO, CA, 94111 |
| NUMBER OF CLAIMS: | 41 |
| EXEMPLARY CLAIM: | 1 |

NUMBER OF DRAWINGS: 7 Drawing Page(s)
LINE COUNT: 4337
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L8 ANSWER 9 OF 27 USPATFULL on STN
TI Human cDNAs and proteins and uses thereof
AB The invention concerns GENSET polynucleotides and polypeptides. Such GENSET products may be used as reagents in forensic analyses, as chromosome markers, as tissue/cell/organelle-specific markers, in the production of expression vectors. In addition, they may be used in screening and diagnosis assays for abnormal GENSET expression and/or biological activity and for screening compounds that may be used in the treatment of GENSET-related disorders.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

| | | | |
|---|---|---|---|
| ACCESSION NUMBER: | 2003:140406 USPATFULL | | |
| TITLE: | Human cDNAs and proteins and uses thereof | | |
| INVENTOR(S): | Bejanin, Stephane, Paris, FRANCE<br>Tanaka, Hiroaki, Antony, FRANCE | | |
| PATENT ASSIGNEE(S): | GENSET, S.A., Paris, FRANCE, 75008 (non-U.S. corporation) | | |
| | NUMBER | KIND | DATE |
| PATENT INFORMATION: | US 2003096247 | A1 | 20030522 |
| APPLICATION INFO.: | US 2001-986 | A1 | 20011114 (10) |
| RELATED APPLN. INFO.: | Division of Ser. No. US 2001-924340, filed on 6 Aug 2001, PENDING | | |
| | NUMBER | DATE | |
| PRIORITY INFORMATION: | WO 2001-IB1715 | 20010806 | |
| | US 2001-305456P | 20010713 (60) | |
| | US 2001-302277P | 20010629 (60) | |
| | US 2001-298698P | 20010615 (60) | |
| | US 2001-293574P | 20010525 (60) | |
| DOCUMENT TYPE: | Utility | | |
| FILE SEGMENT: | APPLICATION | | |
| LEGAL REPRESENTATIVE: | John Lucas, Ph.D., J.D., GENSET CORP., 10665 Sorrento Valley Road, San Diego, CA, 92121-1609 | | |
| NUMBER OF CLAIMS: | 13 | | |
| EXEMPLARY CLAIM: | 1 | | |
| NUMBER OF DRAWINGS: | 4 Drawing Page(s) | | |
| LINE COUNT: | 25656 | | |

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L8 ANSWER 10 OF 27 USPATFULL on STN
TI Human cDNAs and proteins and uses thereof
AB The invention concerns GENSET polynucleotides and polypeptides. Such GENSET products may be used as reagents in forensic analyses, as chromosome markers, as tissue/cell/organelle-specific markers, in the production of expression vectors. In addition, they may be used in screening and diagnosis assays for abnormal GENSET expression and/or biological activity and for screening compounds that may be used in the treatment of GENSET-related disorders.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

| | |
|---|---|
| ACCESSION NUMBER: | 2003:133926 USPATFULL |
| TITLE: | Human cDNAs and proteins and uses thereof |
| INVENTOR(S): | Bejanin, Stephane, Paris, FRANCE<br>Tanaka, Hiroaki, Antony, FRANCE |
| PATENT ASSIGNEE(S): | GENSET, S.A., Paris, FRANCE, 75008 (non-U.S. corporation) |

| | NUMBER | KIND | DATE |
|---|---|---|---|
| PATENT INFORMATION: | US 2003092011 | A1 | 20030515 |
| APPLICATION INFO.: | US 2001-489 | A1 | 20011114 (10) |

RELATED APPLN. INFO.: Division of Ser. No. US 2001-924340, filed on 6 Aug 2001, PENDING

| | NUMBER | DATE |
|---|---|---|
| PRIORITY INFORMATION: | WO 2001-IB1715 | 20010806 |
| | US 2001-305456P | 20010713 (60) |
| | US 2001-302277P | 20010629 (60) |
| | US 2001-298698P | 20010615 (60) |
| | US 2001-293574P | 20010525 (60) |

DOCUMENT TYPE: Utility
FILE SEGMENT: APPLICATION
LEGAL REPRESENTATIVE: John Lucas, Ph.D., J.D., GENSET CORP., 10665 Sorrento Valley Road, San Diego, CA, 92121-1609
NUMBER OF CLAIMS: 13
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 4 Drawing Page(s)
LINE COUNT: 25607
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L8 ANSWER 11 OF 27 USPATFULL on STN

TI Alleviation of the allergenic potential of airborne and contact allergens by thioredoxin

AB Thioredoxin, a small dithiol protein, is a specific reductant for allergenic proteins and particularly allergenic proteins present in pollen and animal and plant sources. All targeted proteins contain disulfide (S--S) bonds that are reduced to the sulfhydryl (SH) level by thioredoxin. The proteins are allergenically active and less digestible in the oxidized (S--S) state. When reduced (SH state), they lose their allergenicity and/or become more digestible. Thioredoxin achieved this reduction when activated (reduced) either by NADPH via NADP-thioredoxin reductase (physiological conditions) or by lipoic acid chemical reductant. Skin tests carried out with sensitized dogs showed that treatment of the pollens with reduced thioredoxin prior to injection eliminated or decreased the allergenicity of the pollen. Studies showed increased digestion of the pollen proteins by pepsin following reduction by thioredoxin. Pollen proteins that have been reduced by thioredoxin are effective and safe immunotherapeutic agents for decreasing or eliminating an animal's allergic reaction that would otherwise occur upon exposure to the non-reduced pollen protein.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.
ACCESSION NUMBER: 2003:115596 USPATFULL
TITLE: Alleviation of the allergenic potential of airborne and contact allergens by thioredoxin
INVENTOR(S): Buchanan, Bob B., Berkeley, CA, United States
del Val, Gregorio, El Cerrito, CA, United States
Lozano, Rosa M., Madrid, SPAIN
Wong, Joshua H., South San Francisco, CA, United States
Yee, Boihon C., Walnut Creek, CA, United States
Frick, Oscar L., San Francisco, CA, United States
PATENT ASSIGNEE(S): Regents of the University of California, Oakland, CA, United States (U.S. corporation)

| | NUMBER | KIND | DATE |
|---|---|---|---|
| PATENT INFORMATION: | US 6555116 | B1 | 20030429 |
| APPLICATION INFO.: | US 1999-238379 | | 19990127 (9) |

RELATED APPLN. INFO.: Continuation-in-part of Ser. No. US 1997-953703, filed on 17 Oct 1997, now patented, Pat. No. US 5952034,

issued on 14 Sep 1999 Continuation-in-part of Ser. No. US 1994-326976, filed on 21 Oct 1994, now patented, Pat. No. US 5792506, issued on 11 Aug 1998 Continuation-in-part of Ser. No. US 211673, now patented, Pat. No. US 6113951, issued on 5 Sep 2000 Continuation-in-part of Ser. No. US 1992-935002, filed on 25 Aug 1992, now abandoned Continuation-in-part of Ser. No. US 1991-776109, filed on 12 Oct 1991, now abandoned

DOCUMENT TYPE: Utility
FILE SEGMENT: GRANTED
PRIMARY EXAMINER: Bugaisky, Gabrielle
LEGAL REPRESENTATIVE: Flehr Hohbach Test Albritton & Herbert LLP, Smith, Karen S.
NUMBER OF CLAIMS: 25
EXEMPLARY CLAIM: 18
NUMBER OF DRAWINGS: 25 Drawing Figure(s); 12 Drawing Page(s)
LINE COUNT: 4670
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L8 ANSWER 12 OF 27 USPATFULL on STN
TI Human cDNAs and proteins and uses thereof
AB The invention concerns GENSET polynucleotides and polypeptides. Such GENSET products may be used as reagents in forensic analyses, as chromosome markers, as tissue/cell/organelle-specific markers, in the production of expression vectors. In addition, they may be used in screening and diagnosis assays for abnormal GENSET expression and/or biological activity and for screening compounds that may be used in the treatment of GENSET-related disorders.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.
ACCESSION NUMBER: 2003:37603 USPATFULL
TITLE: Human cDNAs and proteins and uses thereof
INVENTOR(S): Bejanin, Stephane, Paris, FRANCE
Tanaka, Hiroaki, Antony, FRANCE
PATENT ASSIGNEE(S): GENSET, S.A., Paris, FRANCE, 75008 (non-U.S. corporation)

| | NUMBER | KIND | DATE | |
|---|---|---|---|---|
| PATENT INFORMATION: | US 2003027248 | A1 | 20030206 | |
| APPLICATION INFO.: | US 2001-924340 | A1 | 20010806 | (9) |

| | NUMBER | DATE | |
|---|---|---|---|
| PRIORITY INFORMATION: | US 2001-305456P | 20010713 | (60) |
| | US 2001-302277P | 20010629 | (60) |
| | US 2001-298698P | 20010615 | (60) |
| | US 2001-293574P | 20010525 | (60) |

DOCUMENT TYPE: Utility
FILE SEGMENT: APPLICATION
LEGAL REPRESENTATIVE: GENSET, JOHN LUCAS, PHD, J.D., 10665 SORRENTO VALLEY RD, SAN DIEGO, CA, 92121
NUMBER OF CLAIMS: 13
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 4 Drawing Page(s)
LINE COUNT: 25650
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L8 ANSWER 13 OF 27 USPATFULL on STN
TI Human cDNAs and proteins and uses thereof
AB The invention concerns GENSET polynucleotides and polypeptides. Such GENSET products may be used as reagents in forensic analyses, as chromosome markers, as tissue/cell/organelle-specific markers, in the

production of expression vectors. In addition, they may be used in screening and diagnosis assays for abnormal GENSET expression and/or biological activity and for screening compounds that may be used in the treatment of GENSET-related disorders.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

| | |
|---|---|
| ACCESSION NUMBER: | 2003:37516 USPATFULL |
| TITLE: | Human cDNAs and proteins and uses thereof |
| INVENTOR(S): | Bejanin, Stephane, Paris, FRANCE<br>Tanaka, Hiroaki, Antony, FRANCE |
| PATENT ASSIGNEE(S): | GENSET, S.A., Paris, FRANCE, 75008 (non-U.S. corporation) |

| | NUMBER | KIND | DATE |
|---|---|---|---|
| PATENT INFORMATION: | US 2003027161 | A1 | 20030206 |
| APPLICATION INFO.: | US 2001-992600 | A1 | 20011113 (9) |
| RELATED APPLN. INFO.: | Division of Ser. No. US 2001-924340, filed on 6 Aug 2001, PENDING | | |

| | NUMBER | DATE |
|---|---|---|
| PRIORITY INFORMATION: | WO 2001-IB1715 | 20010806 |
| | US 2001-305456P | 20010713 (60) |
| | US 2001-302277P | 20010629 (60) |
| | US 2001-298698P | 20010615 (60) |
| | US 2001-293574P | 20010525 (60) |

| | |
|---|---|
| DOCUMENT TYPE: | Utility |
| FILE SEGMENT: | APPLICATION |
| LEGAL REPRESENTATIVE: | John Lucas, Ph.D., J.D., GENSET CORP., 10665 Sorrento Valley Road, San Diego, CA, 92121-1609 |
| NUMBER OF CLAIMS: | 13 |
| EXEMPLARY CLAIM: | 1 |
| NUMBER OF DRAWINGS: | 4 Drawing Page(s) |
| LINE COUNT: | 25529 |

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L8 ANSWER 14 OF 27 USPATFULL on STN

TI Methods and compositions for screening for altered cellular phenotypes

AB The invention relates to methods and compositions useful for screening for altered cellular phenotypes using an inducible expression system to enrich for and detect the altered phenotypes and, more particularly, relates to screening libraries of candidate bioactive agents, for example, nucleic acids and peptides, in cells using an regulatable expression system to enrich for a subpopulation of cells having an altered phenotype due to the presence of a candidate bioactive agent.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

| | |
|---|---|
| ACCESSION NUMBER: | 2003:30249 USPATFULL |
| TITLE: | Methods and compositions for screening for altered cellular phenotypes |
| INVENTOR(S): | Lorens, James, Portola Valley, CA, UNITED STATES<br>Kinsella, Todd M., Fayetteville, CA, UNITED STATES<br>Masuda, Esteban, Menlo Park, CA, UNITED STATES<br>Hitoshi, Yasumichi, Mountain view, CA, UNITED STATES<br>Liao, X. Charlene, Palo Alto, CA, UNITED STATES<br>Pearsall, Denise, Belmont, CA, UNITED STATES<br>Friera, Annabelle, South San Francisco, CA, UNITED STATES<br>Chu, Peter, San Francisco, CA, UNITED STATES |

| | NUMBER | KIND | DATE |
|---|---|---|---|
| PATENT INFORMATION: | US 2003022196 | A1 | 20030130 |

APPLICATION INFO.: US 2002-96339 A1 20020308 (10)
RELATED APPLN. INFO.: Continuation-in-part of Ser. No. US 1998-76624, filed on 12 May 1998, PENDING
DOCUMENT TYPE: Utility
FILE SEGMENT: APPLICATION
LEGAL REPRESENTATIVE: DORSEY & WHITNEY LLP, Suite 3400, Four Embarcadero Center, San Francisco, CA, 94111-4187
NUMBER OF CLAIMS: 56
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 50 Drawing Page(s)
LINE COUNT: 5034
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L8 ANSWER 15 OF 27 USPATFULL on STN
TI Synergistic improvements to polynucleotide vaccines
AB The invention features a polynucleotide vaccine modified to enhance expression of the encoded antigen in host cells. The polynucleotide vaccine comprises an antigen-encoding nucleic acid sequence derived from a non-host species of a first phylum or first kingdom, wherein the native signal sequence of the antigen coding sequence is deleted and, optionally, replaced with a signal sequence of a polypeptide of a second phylum or a second kingdom that is functional in the host to be immunized (e.g., a viral signal sequence with a plant antigen-encoding sequence). In one embodiment, the signal sequence is a hemagglutinin A (HA) signal sequence, and the antigen is an allergen (e.g., plant allergen) or from a pathogen (e.g., a bacterium, virus or parasite). The polynucleotide vaccine of the invention provides a synergistic effect with an immunostimulatory sequence (ISS) adjuvant to not only maintain, but to enhance, the immune response to the encoded antigen.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.
ACCESSION NUMBER: 2002:259405 USPATFULL
TITLE: Synergistic improvements to polynucleotide vaccines
INVENTOR(S): Raz, Eyal, Del Mar, CA, UNITED STATES
Takabayashi, Kenji, San Diego, CA, UNITED STATES
Nguyen, Minh-Duc, Oceanside, CA, UNITED STATES

| | NUMBER | KIND | DATE | |
|---|---|---|---|---|
| PATENT INFORMATION: | US 2002142978 | A1 | 20021003 | |
| APPLICATION INFO.: | US 2001-828505 | A1 | 20010406 | (9) |

| | NUMBER | DATE |
|---|---|---|
| PRIORITY INFORMATION: | US 2000-195890P | 20000407 (60) |

DOCUMENT TYPE: Utility
FILE SEGMENT: APPLICATION
LEGAL REPRESENTATIVE: Carol L. Francis, BOZICEVIC, FIELD & FRANCIS LLP, Suite 200, 200 Middlefield Road, Menlo Park, CA, 94025
NUMBER OF CLAIMS: 32
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 7 Drawing Page(s)
LINE COUNT: 2072
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L8 ANSWER 16 OF 27 USPATFULL on STN
TI Stabilization of hypoallergenic, hyperdigestible previously reduced proteins
AB Disulfide proteins showed mitigated allergenicity and increased digestibility by pepsin following reduction by thioredoxin. The sulfhydryl groups newly formed on reduction by thioredoxin (at 4.degree. C.) or dithiothreitol (DTT) (at 55.degree. C.) were blocked with a physiological disulfide, such as cystamine or oxidized glutathione (GSSG) to obtain stable forms of the disarmed allergen. When derivatized

with cystamine, BLG was separated from its oxidized and reduced forms on non-reducing SDS-PAGE and appeared to lack sulfhydryl groups. Although less effective GSSG, gave similar results. Allergenicity of the two derivatives was compared with that of the oxidized, reduced and reoxidized forms of BLG by skin testing dogs from a colony sensitized to cow's milk. Both the cystamine and GSSG derivatized BLG showed decreased allergenicity and increased sensitivity to pepsin as compared to controls. The reoxidized form resembled the derivatives in having lower allergenicity. The thioredoxin- and DTT-reduced forms showed hypoallergenic, hyperdigestible properties, most effectively when the reduced proteins were heated at 55.degree. C. Whole milk subjected to these procedures showed results similar to those obtained with pure BLG. Other proteins are similarly stabilized. Stable forms of such disarmed, hypoallergenic and hyperdigestible disulfide protein allergens or just hypoallergenic or just hyperdigestible protein allergens are useful in foods as well as clinical preparations.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.
ACCESSION NUMBER: 2002:185367 USPATFULL
TITLE: Stabilization of hypoallergenic, hyperdigestible previously reduced proteins
INVENTOR(S): Buchanan, Bob B., Berkeley, CA, UNITED STATES
Morigasaki, Susumu, Berkeley, CA, UNITED STATES
Val, Gregorio del, San Diego, CA, UNITED STATES
Frick, Oscar L., San Francisco, CA, UNITED STATES

| | NUMBER | KIND | DATE |
|---|---|---|---|
| PATENT INFORMATION: | US 2002098277 | A1 | 20020725 |
| APPLICATION INFO.: | US 2001-779375 | A1 | 20010207 (9) |

RELATED APPLN. INFO.: Continuation-in-part of Ser. No. US 1999-238379, filed on 27 Jan 1999, UNKNOWN
DOCUMENT TYPE: Utility
FILE SEGMENT: APPLICATION
LEGAL REPRESENTATIVE: FLEHR HOHBACH TEST, ALBRITTON & HERBERT LLP, Suite 3400, Four Embarcadero Center, San Francisco, CA, 94111-4187
NUMBER OF CLAIMS: 26
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 27 Drawing Page(s)
LINE COUNT: 5501
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L8 ANSWER 17 OF 27 USPATFULL on STN
TI Vaccines comprising oil bodies
AB The present invention provides novel adjuvants which comprise oil bodies. The invention also provides vaccine formulations comprising oil bodies and an antigen and methods for preparing the vaccines and the use of the vaccines to elicit an immune response.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.
ACCESSION NUMBER: 2002:140865 USPATFULL
TITLE: Vaccines comprising oil bodies
INVENTOR(S): Deckers, Harm M., Alberta, CANADA
Rooijen, Gijs Van, Alberta, CANADA
Boothe, Joseph, Alberta, CANADA
Goll, Janis, Alberta, CANADA
Moloney, Maurice M., Alberta, CANADA
Schryvers, Anthony B., Alberta, CANADA
Alcantara, Joenel, Alberta, CANADA
Hutchins, Wendy A., Alberta, CANADA

| | NUMBER | KIND | DATE |
|---|---|---|---|

PATENT INFORMATION: US 2002071846 A1 20020613
APPLICATION INFO.: US 2001-880901 A1 20010615 (9)
RELATED APPLN. INFO.: Continuation-in-part of Ser. No. US 2000-577147, filed on 24 May 2000, PENDING Continuation-in-part of Ser. No. US 1999-448600, filed on 24 Nov 1999, PATENTED Continuation-in-part of Ser. No. US 1998-84777, filed on 27 May 1998, PATENTED

| | NUMBER | DATE |
|---|---|---|
| PRIORITY INFORMATION: | US 1998-75863P | 19980225 (60) |
| | US 1998-75864P | 19980225 (60) |
| | US 1997-47779P | 19970528 (60) |
| | US 1997-47753P | 19970527 (60) |
| | US 2000-212130P | 20000616 (60) |

DOCUMENT TYPE: Utility
FILE SEGMENT: APPLICATION
LEGAL REPRESENTATIVE: BURNS DOANE SWECKER & MATHIS L L P, POST OFFICE BOX 1404, ALEXANDRIA, VA, 22313-1404
NUMBER OF CLAIMS: 27
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 10 Drawing Page(s)
LINE COUNT: 2348
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L8 ANSWER 18 OF 27 USPATFULL on STN
TI TRUNCATED CRAFI INHIBITS CD40 SIGNALING
AB Overexpression of a CRAF1 (CD40 receptor-associated factor 1) gene truncated by 323 to about 414 amino acids at the amino inhibits CD40-mediated cell activation, and is used to treat conditions characterized by an unwanted level of CD40-mediated intracellular signaling.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.
ACCESSION NUMBER: 2002:54363 USPATFULL
TITLE: TRUNCATED CRAFI INHIBITS CD40 SIGNALING
INVENTOR(S): BALTIMORE, DAVID, BOSTON, MA, UNITED STATES
CHENG, GENHONG, LOS ANGELES, CA, UNITED STATES
YE, ZHENG-SHENG, NEW YORK, NY, UNITED STATES
LEDERMAN, SETH, NEW YORK, NY, UNITED STATES
CLEARY, AILEEN, NEW YORK, NY, UNITED STATES

| | NUMBER | KIND | DATE |
|---|---|---|---|
| PATENT INFORMATION: | US 2002031522 | A1 | 20020314 |
| APPLICATION INFO.: | US 1997-813323 | A1 | 19970310 (8) |

| | NUMBER | DATE |
|---|---|---|
| PRIORITY INFORMATION: | US 1996-13199P | 19960311 (60) |

DOCUMENT TYPE: Utility
FILE SEGMENT: APPLICATION
LEGAL REPRESENTATIVE: JOHN P WHITE, COOPER AND DUNHAM, 1185 AVENUE OF THE AMERICAS, NEW YORK, NY, 10036
NUMBER OF CLAIMS: 91
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 11 Drawing Page(s)
LINE COUNT: 1555
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L8 ANSWER 19 OF 27 USPATFULL on STN
TI Heat shock fusion-based vaccine system
AB Disclosed are epitope-containing heat shock fusion proteins, DNA constructs encoding such fusion proteins, and methods of use. More

specifically, disclosed are ubiquitin fusion proteins comprising ubiquitin fused to a plurality of identical or non-identical epitopes at specified locations.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

| | |
|---|---|
| ACCESSION NUMBER: | 2001:208483 USPATFULL |
| TITLE: | Heat shock fusion-based vaccine system |
| INVENTOR(S): | Kenten, John H., Boyds, MD, United States<br>Tramontano, Alfonso, Rockville, MD, United States<br>Pilon, Aprile L., Gaithersburg, MD, United States<br>Lohnas, Gerald L., Catonsville, MD, United States<br>Roberts, Steven F., Bethesda, MD, United States |
| PATENT ASSIGNEE(S): | Proteinix Company, Gaithersburg, MD, United States (U.S. corporation) |

| | NUMBER | KIND | DATE | |
|---|---|---|---|---|
| PATENT INFORMATION: | US 6319503 | B1 | 20011120 | |
| APPLICATION INFO.: | US 1998-26276 | | 19980219 | (9) |

| | |
|---|---|
| DOCUMENT TYPE: | Utility |
| FILE SEGMENT: | GRANTED |
| PRIMARY EXAMINER: | Saoud, Christine J. |
| ASSISTANT EXAMINER: | Hamud, Fozia |
| LEGAL REPRESENTATIVE: | Farrell, Kevin M. |
| NUMBER OF CLAIMS: | 17 |
| EXEMPLARY CLAIM: | 1 |
| LINE COUNT: | 1494 |

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L8 ANSWER 20 OF 27 USPATFULL on STN

TI Multivalent compounds for crosslinking receptors and uses thereof

AB Synthetic crosslinking homobivalent and heterobivalent compounds have been designed and developed. These compounds are low in molecular weight, have antagonistic or agonistic activity, and induce the association between two identical or similar natural receptors (homobivalent compounds) or induce the association between two different natural receptors (heterobivalent compounds).

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

| | |
|---|---|
| ACCESSION NUMBER: | 1999:124868 USPATFULL |
| TITLE: | Multivalent compounds for crosslinking receptors and uses thereof |
| INVENTOR(S): | Bachovchin, William W., Melrose, MA, United States |
| PATENT ASSIGNEE(S): | Trustees of Tufts College, Medford, MA, United States (U.S. corporation) |

| | NUMBER | KIND | DATE | |
|---|---|---|---|---|
| PATENT INFORMATION: | US 5965532 | | 19991012 | |
| APPLICATION INFO.: | US 1997-837305 | | 19970411 | (8) |

| | |
|---|---|
| RELATED APPLN. INFO.: | Continuation-in-part of Ser. No. US 1996-671756, filed on 28 Jun 1996, now abandoned |
| DOCUMENT TYPE: | Utility |
| FILE SEGMENT: | Granted |
| PRIMARY EXAMINER: | Huff, Sheela |
| LEGAL REPRESENTATIVE: | Wolf, Greenfield & Sacks, P.C. |
| NUMBER OF CLAIMS: | 11 |
| EXEMPLARY CLAIM: | 1 |
| NUMBER OF DRAWINGS: | 38 Drawing Figure(s); 20 Drawing Page(s) |
| LINE COUNT: | 3884 |

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L8 ANSWER 21 OF 27 USPATFULL on STN

TI Increasing the digestibility of food proteins by thioredoxin reduction

AB Thioredoxin, a small dithiol protein, is a specific reductant for major food proteins, allergenic proteins and particularly allergenic proteins present in widely used foods from animal and plant sources. All targeted proteins contain disulfide (S--S) bonds that are reduced to the sulfhydryl (SH) level by thioredoxin. The proteins are allergenically active and less digestible in the oxidized (S--S) state. When reduced (SH state), they lose their allergenicity and/or become more digestible. Thioredoxin achieved this reduction when activated (reduced) either by NADPH via NADP-thioredoxin reductase (physiological conditions) or by dithiothreitol, a chemical reductant. Skin tests and feeding experiments carried out with sensitized dogs showed that treatment of the food with reduced thioredoxin prior to ingestion eliminated or decreased the allergenicity of the food. Studies showed increased digestion of food and food proteins by pepsin and trypsin following reduction by thioredoxin.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

| | |
|---|---|
| ACCESSION NUMBER: | 1999:110024 USPATFULL |
| TITLE: | Increasing the digestibility of food proteins by thioredoxin reduction |
| INVENTOR(S): | Buchanan, Bob B., Berkeley, CA, United States<br>del Val, Gregorio, Saint-Aubin/NE, Switzerland<br>Lozano, Rosa M., Madrid, Spain<br>Jiao, Jin-an, Ft. Lauderdale, FL, United States<br>Wong, Joshua H., South San Francisco, CA, United States<br>Yee, Boihon C., Walnut Creek, CA, United States |
| PATENT ASSIGNEE(S): | The Regents of the University of California, Oakland, CA, United States (U.S. corporation) |

| | NUMBER | KIND | DATE |
|---|---|---|---|
| PATENT INFORMATION: | US 5952034 | | 19990914 |
| APPLICATION INFO.: | US 1997-953703 | | 19971017 (8) |

| | |
|---|---|
| RELATED APPLN. INFO.: | Continuation-in-part of Ser. No. US 1994-326976, filed on 21 Oct 1994, now patented, Pat. No. US 5792506 which is a continuation-in-part of Ser. No. US 1994-211673, filed on 12 Apr 1994 which is a continuation-in-part of Ser. No. US 1992-935002, filed on 25 Aug 1992, now abandoned which is a continuation-in-part of Ser. No. US 1991-776109, filed on 12 Oct 1991, now abandoned |
| DOCUMENT TYPE: | Utility |
| FILE SEGMENT: | Granted |
| PRIMARY EXAMINER: | Hendricks, Keith D. |
| LEGAL REPRESENTATIVE: | Smith, Karen S.Flehr Hohbach Test Albritton & Herbert LLP |
| NUMBER OF CLAIMS: | 6 |
| EXEMPLARY CLAIM: | 1 |
| NUMBER OF DRAWINGS: | 18 Drawing Figure(s); 9 Drawing Page(s) |
| LINE COUNT: | 4164 |

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L8 ANSWER 22 OF 27 USPATFULL on STN

TI Isolation and characterization of allergen-binding cells for diagnosis of hypersensitivity

AB Methods and compositions are provided for the diagnosis of allergen hypersensitivity in a patient. Rare, allergen-specific cells are enriched from a complex cell population, e.g. a patient blood sample. The percentage of blood cells that bind to a particular allergen is less than 0.01%. The allergen-specific cell population is enriched by magnetic cell sorting. In normal blood, the allergen-binding cells are primarily B-cells expressing CD19 and CD21. In blood from allergic patients, an additional population of effector cells, e.g. basophilic granulocytes is labeled by the allergen.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.
ACCESSION NUMBER: 1999:72511 USPATFULL
TITLE: Isolation and characterization of allergen-binding cells for diagnosis of hypersensitivity
INVENTOR(S): Irsch, Johannes, Cologne, Germany, Federal Republic of
Miltenyi, Stefan, Bergisch Gladbach, Germany, Federal Republic of
Radbruch, Andreas, Berlin, Germany, Federal Republic of
PATENT ASSIGNEE(S): Miltenyi Biotec GmbH, Bergisch Gladbach, Germany, Federal Republic of (non-U.S. corporation)

| | NUMBER | KIND | DATE |
|---|---|---|---|
| PATENT INFORMATION: | US 5916818 | | 19990629 |
| APPLICATION INFO.: | US 1998-37126 | | 19980309 (9) |

RELATED APPLN. INFO.: Continuation of Ser. No. US 1996-660035, filed on 6 Jun 1996, now patented, Pat. No. US 5786161
DOCUMENT TYPE: Utility
FILE SEGMENT: Granted
PRIMARY EXAMINER: Scheiner, Toni R.
LEGAL REPRESENTATIVE: Cooley Godward LLP
NUMBER OF CLAIMS: 6
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 9 Drawing Figure(s); 4 Drawing Page(s)
LINE COUNT: 752
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L8 ANSWER 23 OF 27 USPATFULL on STN
TI Methods and peptides for the treatment of non-IgE-mediated diseases
AB Methods and compositions for the treatment of non-Ige-mediated inflammatory response or disease conditions are described.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.
ACCESSION NUMBER: 1999:67252 USPATFULL
TITLE: Methods and peptides for the treatment of non-IgE-mediated diseases
INVENTOR(S): Hahn, Gary S., Cardiff by the Sea, CA, United States
PATENT ASSIGNEE(S): Dura Pharmaceuticals, Inc., San Diego, CA, United States (U.S. corporation)

| | NUMBER | KIND | DATE |
|---|---|---|---|
| PATENT INFORMATION: | US 5912233 | | 19990615 |
| APPLICATION INFO.: | US 1995-462304 | | 19950605 (8) |

RELATED APPLN. INFO.: Continuation-in-part of Ser. No. US 1992-942671, filed on 8 Sep 1992 which is a continuation of Ser. No. US 1992-878867, filed on 5 May 1992 which is a continuation-in-part of Ser. No. US 411489
DOCUMENT TYPE: Utility
FILE SEGMENT: Granted
PRIMARY EXAMINER: Davenport, Avis M.
LEGAL REPRESENTATIVE: Lyon & Lyon LLP
NUMBER OF CLAIMS: 16
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 8 Drawing Figure(s); 4 Drawing Page(s)
LINE COUNT: 1000
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L8 ANSWER 24 OF 27 USPATFULL on STN
TI Isolation and characterization of allergen-binding cells for diagnosis of hypersensitivity
AB Methods and compositions are provided for the diagnosis of allergen hypersensitivity in a patient. Rare, allergen-specific cells are enriched from a complex cell population, e.g. a patient blood sample.

The percentage of blood cells that bind to a particular allergen is less than 0.01%. The allergen-specific cell population is enriched by magnetic cell sorting. In normal blood, the allergen-binding cells are primarily B-cells expressing CD19 and CD21. In blood from allergic patients, an additional population of effector cells, e.g. basophilic granulocytes is labeled by the allergen.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

| | |
|---|---|
| ACCESSION NUMBER: | 1998:88655 USPATFULL |
| TITLE: | Isolation and characterization of allergen-binding cells for diagnosis of hypersensitivity |
| INVENTOR(S): | Irsch, Johannes, Cologne, Germany, Federal Republic of<br>Miltenyi, Stefan, Bergisch Gladbach, Germany, Federal Republic of<br>Radbruch, Andreas, Cologne, Germany, Federal Republic of |
| PATENT ASSIGNEE(S): | Miltenyi Biotec. GmbH, Bergisch Gladbach, Germany, Federal Republic of (non-U.S. corporation) |

| | NUMBER | KIND | DATE | |
|---|---|---|---|---|
| PATENT INFORMATION: | US 5786161 | | 19980728 | |
| APPLICATION INFO.: | US 1996-660035 | | 19960606 | (8) |

| | |
|---|---|
| DOCUMENT TYPE: | Utility |
| FILE SEGMENT: | Granted |
| PRIMARY EXAMINER: | Scheiner, Toni R. |
| LEGAL REPRESENTATIVE: | Cooley Godward LLP |
| NUMBER OF CLAIMS: | 18 |
| EXEMPLARY CLAIM: | 18 |
| NUMBER OF DRAWINGS: | 9 Drawing Figure(s); 4 Drawing Page(s) |
| LINE COUNT: | 836 |

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L8 ANSWER 25 OF 27 USPATFULL on STN
TI Methods for the treatment of non-IgE-mediated diseases
AB Methods and compositions for the treatment of non-Ige-mediated inflammatory response or disease conditions are described.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

| | |
|---|---|
| ACCESSION NUMBER: | 95:103492 USPATFULL |
| TITLE: | Methods for the treatment of non-IgE-mediated diseases |
| INVENTOR(S): | Hahn, Gary S., Cardiff by the Sea, CA, United States |
| PATENT ASSIGNEE(S): | Dura Pharmaceuticals, Inc., San Diego, CA, United States (U.S. corporation) |

| | NUMBER | KIND | DATE | |
|---|---|---|---|---|
| PATENT INFORMATION: | US 5468730 | | 19951121 | |
| APPLICATION INFO.: | US 1992-942671 | | 19920908 | (7) |

| | |
|---|---|
| DISCLAIMER DATE: | 20081029 |
| RELATED APPLN. INFO.: | Continuation of Ser. No. US 1992-878867, filed on 5 May 1992, now abandoned which is a continuation-in-part of Ser. No. US 1989-411489, filed on 23 Nov 1989, now abandoned |
| DOCUMENT TYPE: | Utility |
| FILE SEGMENT: | Granted |
| PRIMARY EXAMINER: | Warden, Jill |
| ASSISTANT EXAMINER: | Davenport, A. M. |
| LEGAL REPRESENTATIVE: | Lyon & Lyon |
| NUMBER OF CLAIMS: | 6 |
| EXEMPLARY CLAIM: | 1 |
| NUMBER OF DRAWINGS: | 8 Drawing Figure(s); 4 Drawing Page(s) |
| LINE COUNT: | 946 |

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L8 ANSWER 26 OF 27 USPATFULL on STN
TI Methods and compositions for the treatment of non-IgE-mediated diseases
AB Methods and compositions for the treatment of non-IgE-mediated inflammatory disease conditions utilizing the peptide Asp-Ser-Asp-Pro-Arg, or derivative thereof are disclosed.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

| | | | |
|---|---|---|---|
| ACCESSION NUMBER: | 91:89037 USPATFULL | | |
| TITLE: | Methods and compositions for the treatment of non-IgE-mediated diseases | | |
| INVENTOR(S): | Hahn, Gary S., Cardiff by the Sea, CA, United States | | |
| PATENT ASSIGNEE(S): | Immunetech Pharmaceuticals, San Diego, CA, United States (U.S. corporation) | | |
| | NUMBER | KIND | DATE |
| PATENT INFORMATION: | US 5061692 | | 19911029 |
| | WO 8804177 | | 19880716 |
| APPLICATION INFO.: | US 1989-382623 | | 19891123 (7) |
| | WO 1987-US3222 | | 19871209 |
| | | | 19891123 PCT 371 date |
| | | | 19891123 PCT 102(e) date |
| RELATED APPLN. INFO.: | Continuation-in-part of Ser. No. US 1986-939927, filed on 9 Dec 1986, now patented, Pat. No. US 4816449 which is a continuation-in-part of Ser. No. US 1986-899891, filed on 25 Aug 1986, now abandoned which is a continuation of Ser. No. US 1986-824945, filed on 3 Feb 1986, now patented, Pat. No. US 4628045 which is a continuation of Ser. No. US 1985-746175, filed on 18 Jun 1985, now abandoned which is a continuation-in-part of Ser. No. US 1983-522601, filed on 12 Aug 1983, now abandoned | | |
| DOCUMENT TYPE: | Utility | | |
| FILE SEGMENT: | Granted | | |
| PRIMARY EXAMINER: | Lee, Lester L. | | |
| LEGAL REPRESENTATIVE: | Lyon & Lyon | | |
| NUMBER OF CLAIMS: | 10 | | |
| EXEMPLARY CLAIM: | 1 | | |
| LINE COUNT: | 558 | | |

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L8 ANSWER 27 OF 27 WPIDS COPYRIGHT 2003 THOMSON DERWENT on STN
TI New vaccine comprising **allergy** peptides linked by an inert carrier, useful for boosting an anti-**allergy** immune response in an individual susceptible to an **allergic response**.
AN 2001-091150 [10] WPIDS
AB WO 200074716 A UPAB: 20010220
NOVELTY - A composition comprising **allergy** peptides linked by an inert carrier is new. The **allergy** peptides are derived from Immunoglobulin E (IgE) or IgE receptor. The inert carrier does not contain a peptidic T-cell helper epitope, which is capable of binding to an MHC (major histocompatability) molecule and is capable of stimulating T-cell proliferation.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:

(1) a method of boosting an anti-**allergy** immune response by administering the composition as a vaccine to an individual susceptible to an **allergic response**, where the immune system of the individual has previously been primed with the composition comprising the **allergy** peptide and a carrier comprising a peptidic T-cell helper epitope;

(2) a method for producing an **allergy** vaccine comprising manufacturing the composition and formulating the composition with an

adjuvant; and

(3) a method of inducing and maintaining an anti-**allergy** effective immune response comprising:

(a) administering to an individual the composition comprising an **allergy** peptide conjugated to a T-cell epitope containing carrier; and

(b) a subsequent administration to the individual of the composition comprising the **allergy** peptide in the absence of a peptidic T-cell epitope containing carrier.

ACTIVITY - Antiallergic.

MECHANISM OF ACTION - Vaccine.

Female (B6 multiply BALB/c)F1 mice were primed with: (i) a protein D-stanworth decapeptide (KTKGSGFFVF) conjugate formulated in oil in water emulsion adjuvant with 3D-MPL and QS21, or (ii) decapeptide PS6B polysaccharide conjugate formulated in oil in water emulsion adjuvant with 3D-MPL and QS21. All mice were boosted 14 months after the second priming dose. The results showed that Deca-peptide-PS did give a boost when formulated with an adjuvant in the absence of peptide T-cell helper epitopes. In the deca-PS6B group, 9/10 mice had a 4-fold increase in anti-deca titer after boosting in comparison to the pre-boost titers (4/10 had a 16-fold increase). Furthermore, 10/10 of mice in group (i) (deca-protein D conjugate) had at least a 4-fold increase (only 2/10 showed a 16-fold increase).

USE - The composition is useful as a vaccine or for manufacturing a medicament for the prophylaxis or treatment of **allergy**. In particular for boosting an anti-**allergy** immune response in an individual susceptible to an **allergic response**.

Dwg.0/0

ACCESSION NUMBER: 2001-091150 [10] WPIDS
DOC. NO. CPI: C2001-026765
TITLE: New vaccine comprising **allergy** peptides linked by an inert carrier, useful for boosting an anti-**allergy** immune response in an individual susceptible to an **allergic response**.
DERWENT CLASS: B04 D16
INVENTOR(S): PRIEELS, J
PATENT ASSIGNEE(S): (SMIK) SMITHKLINE BEECHAM BIOLOGICALS
COUNTRY COUNT: 93
PATENT INFORMATION:

| PATENT NO | KIND | DATE | WEEK | LA | PG |
|---|---|---|---|---|---|
| WO 2000074716 | A2 | 20001214 | (200110)* | EN | 26 |

RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ NL OA PT SD SE SL SZ TZ UG ZW

W: AE AG AL AM AT AU AZ BA BB BG BR BY CA CH CN CR CU CZ DE DK DM DZ EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW

AU 2000058116 A 20001228 (200119)

APPLICATION DETAILS:

| PATENT NO | KIND | APPLICATION | DATE |
|---|---|---|---|
| WO 2000074716 | A2 | WO 2000-EP5164 | 20000606 |
| AU 2000058116 | A | AU 2000-58116 | 20000606 |

FILING DETAILS:

| PATENT NO | KIND | | PATENT NO |
|---|---|---|---|
| AU 2000058116 | A | Based on | WO 2000074716 |